January 2010

Unique Features of PTB RRMs: Insight into Protein Motions and RNA Binding

Caroline Maynard
Washington University in St. Louis

WASHINGTON UNIVERSITY IN ST. LOUIS

Division of Biology and Biomedical Sciences

Molecular Biophysics

Dissertation Examination Committee:
Kathleen B. Hall (chair)
Carl Frieden
Nathan Baker
Peter Chivers
Katherine Henzler-Wildman
Garland Marshall

UNIQUE FEATURES OF POLYPYRIMIDINE TRACT BINDING PROTEIN (PTB) RNA RECOGNITION MOTIFS (RRMS): INSIGHT INTO PROTEIN MOTIONS AND RNA BINDING

by

Caroline Marie Maynard

A dissertation presented to the
Graduate School of Arts and Sciences
of Washington University in
partial fulfillment of the
requirements for the degree
of Doctor of Philosophy

August 2010

Saint Louis, Missouri

Abstract of the Dissertation

Polypyrimidine tract binding protein (PTB) is a highly conserved RNA binding protein comprised of four RRMs (RNA recognition motifs). RRMs are extremely prevalent in all kingdoms of life, and have been very well characterized in terms of structure and RNA binding properties. However, all four RRMs of PTB exhibit unique features that can be exploited to learn more about the RNA selection and binding strategy of PTB, as well more general features involving structure/function relationships and protein unfolding mechanisms.

PTB participates in a variety of functions in eukaryotic cells, including alternative splicing, mRNA stabilization, and internal ribosomal entry site (IRES) mediated translation initiation. Its mechanism of RNA recognition is determined in part by the novel geometry of its two C-terminal RNA Recognition Motifs (RRM3 and RRM4), which interact with each other to form a stable complex (PTB1:34). This complex itself is unusual among RRMs, suggesting that it performs a specific function for the protein. In order to understand the advantage it provides to PTB, the fundamental properties of PTB1:34 are examined here as a comparative study of the complex and its two constituent RRMs. Both RRM3 and RRM4 adopt folded, and reasonably stable structures, yet the RNA binding properties of the domains differ dramatically. RRM4 does not bind to RNA, and although RRM3 binds to polypyrimidine tracts, its affinity is significantly weaker than that of PTB1:34. ^{15}N-NMR relaxation experiments show that the interaction between RRM3 and RRM4 induces microsecond motions throughout PTB1:34 and forms a unique RNA binding platform. The motions could be important for

RNA selection based on secondary structure, part of an RNA binding mechanism, entropic compensation for formation of PTB1:34, or a mechanism of allosteric communication between binding sites of the RRMs. A mutant protein was designed to address the contribution of the motions to protein function.

PTB RRM2 and RRM3 are structurally unique in that they both have a C-terminal extension that adds a fifth β-strand to the canonical four stranded β-sheet, connected to β-strand four by a flexible linker. This extension both extends and occludes the putative RNA binding surface. Other RRM extensions have been reported, and appear to influence protein function through a variety of mechanisms including direct interactions with RNA, participation in protein-protein interactions, or stabilization of the RRM core domain. Studies using a truncated form of RRM3 that lacks the extension show that, in this case, the contribution to protein function is likely due to direct RNA contacts.

Finally, PTB RRM4 has a unique chemical melting profile that may be useful for investigating protein unfolding transitions. A tryptophan mutant was engineered to facilitate fluorescence studies, and the protein was found to be a natural 'missing link' between two-state and downhill folders.

Acknowledgements

This work was supported by:

NIH grant GM077231 (K.B.H)

NIH training grant T32 GM008492

Sigma-Cori predoctoral fellowship (C.M.M.)

Special thanks to my many teachers, friends and family.
Your support has been invaluable.

Table of Contents

List of Tables and Figures

Chapter 1

Introduction to the Dissertation

Today, centuries of research and development of technology culminate to allow medical science to function at the molecular level. We now understand that most biological processes depend on the interaction between specific molecules within the cell. Proteins have long been known as the 'workhorses' of the cell, and are responsible for vast and very different biological jobs including catalysis, providing scaffolding, and myriad interactions with other cellular components. Yet, a predictive understanding of how these tiny machines work is still lacking.

Traditional molecular biology dogma asserts that the primary sequence of a protein determines its three-dimensional structure, and that a protein's three-dimensional structure determines its function. These basic premises are limiting and outdated in terms of understanding molecular mechanisms, and should be challenged. I am most interested in exploring biological systems where the above assertions are not necessarily true. In addition, it is my opinion that inherent motions of biological molecules are under-appreciated and that this hinders progress towards a predictive understanding of protein function. Although much work has been poured into learning the mechanisms of protein function, the understanding of protein motions in these processes remains limited. It is merely common sense to surmise that any machine, including protein machines, require moving parts, and recent scientific advances have shown that motions are important for biological molecules. I have spent the last several years exploring a system that can be used to address the above issues in the hopes that the work described herein may

contribute to freeing the problem of delineating molecular mechanisms from the above constraints, and may provide fresh avenues for original thought on protein function.

When I discuss my work with non-scientists, I frequently use the following analogy. Suppose you had never seen a bicycle, and someone gave you a photograph of a bicycle and told you to figure out how it works, so you could build another one. This task would be difficult to impossible. Now imagine you had access to a movie of the bicycle working, with moving parts in motion. The mechanism of the bicycle in this case is obvious. Understanding molecular mechanisms is more complicated, and as such, only stands to gain more benefit from developing 'movies' of the molecules. This illustrates why, even after decades of research, and huge advances in structural biology, the field is still hard pressed to define structure-function relationships from static three dimensional structures. Yet many trained scientists reject the importance of molecular motions. With this work, I hope to provide one more piece to the puzzle of understanding how proteins function, and show, in the system described here, that motions are an important aspect of a protein's inherent nature.

Polypyrimidine tract binding protein

The polypyrimidine tract binding protein (PTB) was chosen for these studies for a number of reasons. First, PTB is found in all eukaryotic tissues and developmental stages, and has been found to be important for a wide range of processes that take place in both the nucleus and cytoplasm including alternative spicing, mRNA stabilization, polyadenylation and non-traditional translation initiation. In order to participate in such

far-reaching processes, PTB must be able to interact with specific RNAs as well as other proteins in a tightly controlled manner. Thus, understanding how this versatile protein accomplishes such a wide range of tasks is not only requisite for understanding the processes themselves, but may provide a therapeutic target or strategy.

Moreover, PTB has already been studied in great detail. The structure has been solved (Conte et al., 2000; Simpson et al, 2004; Oberstrass et al., 2005), and extensive binding studies have been completed and binding specificities reported for both protein (Clerte and Hall, 2006; Clerte and Hall, 2009; Yuan et al., 2002; Simpson et al., 2004; Oberstrass et al., 2005) and RNA (Perez et al. 1997), allowing further research to focus on some of the finer details of the system. Finally, PTB is comprised entirely of four RNA recognition motifs (RRMs) separated by conserved linkers. RRMs are the most common eukaryotic RNA binding domain, and the structure and function of many RRMs are well established. Thus, the known structure and binding characteristics provide a context for interpretation of detailed thermodynamic and motional analyses, and the fact that the protein is comprised of RRMs allows comparison to a well studied family of proteins enabling broader questions pertaining to protein structure and function to be addressed.

PTB is involved in a wide range of cell processes

PTB has been found to be important for a multitude of cellular processes both in the nucleus and cytoplasm. Participation in these regulatory events can occur through localization, binding competition or post-transcriptional modifications. PTB is most well

known for its role in alternative splicing, and is usually inhibitory (Valcarcel and Gebauer, 1997), but has also been found to enhance some alternative splicing events (Lou et al. 1999, Shen et al. 2004). Interestingly, PTB is able to control its own expression levels by binding to its mRNA through an alternative splicing mechanism that results in nonsense mediated decay of the resulting frame shifted mRNA (Wollerton et al., 2004). Other alterative splicing events are capable of producing three isoforms, PTB1, PTB2 and PTB4, which differ only in the length of the linker between RRM2 and RRM3 (Ghetti, et al., 1992). PTB1 is by far the best characterized, and is the form used here.

PTB also participates in other cellular regulation events including RNA 3'-end processing, especially mRNA polyadenylation (Castelo-Branco et al. 2004) and typically appears to promote polyadenylation in concert with splicing repression (Lou et al., 1999; LeSommer et al., 2005), suggesting a dual role for PTB in some systems. PTB has also been reported to stabilize mRNA by binding to UTRs (either 3' or 5') for insulin (Tillmar et al., 2002), vascular endothelial growth factor (Coles et al., 2004), CD154 (Hamilton et al., 2003), inducible nitric oxide sythase (Pautz et al., 2006), and phosphoglycerate kinase 2 (Xu and Hecht, 2007).

PTB contains both an NLS (nuclear localization signal) and an NES (nuclear export signal), allowing it to shuttle efficiently between the nucleus and cytoplasm (Ghetti et al., 1992; Li and Yen, 2002). Translocation to the cytoplasm is usually triggered by cell stress, such as viral infection, apoptosis, and exposure to genotoxic substances. In addition, PTB has been implicated in RNA localization processes (Cote et

al., 1999). PTB interacts with α-actin mRNA, where it localizes at neurite growth terminals, important for cell motility and neuronal axon growth (Ma et al., 2007), and also has been found to be a nuclear export factor for hepatitis B virus (HBV) RNA (Zang et al., 2001).

The participation of PTB in internal ribosome entry site (IRES) translation initiation has also been well documented. Under conditions of cell stress, when cap-dependent translation is inhibited (Bushell et al., 2006), or for some viruses, such as picornaviruses, that lack a 5'-mRNA cap structure, this is an important process for protein production. In many systems, including many picornaviral and flaviviral entry sites, PTB is required for translation initiation (Hellen et al., 1993; Pilipenko et al., 2000; Gosert et al., 2000), but has also been found to repress IRES translation (Cornelis et al., 2005). In some cases, PTB appears to function as an RNA chaperone, but has also been shown to have a role in ribosome recruitment (Song et al., 2005). Furthermore, PTB binds to hepatitis C virus IRESs (Gontarek et al., 1999), and co-localizes with the HCV replication complex, but is not required for IRES activity (Mitchell, et al., 2005).

Finally, truncated versions of PTB have been found to be biologically significant units. An alternative splicing event that produces a protein containing only the two C-terminal RRMs of PTB (PTB1:34) was identified in activated T lymphocytes (Hamilton et al., 2003). This protein binds to the CD154 3'-UTR and modulates protein expression. Furthermore, both poliovirus (PV) and hepatitis A virus (HAV) proteinase 3C cleaves PTB between RRM2 and RRM3 producing PTB1:34 (Back et al., 2002; Kanda et al.,

2009). Thus, investigation of both full length PTB and PTB1:34 is of significant biological interest.

PTB has some unique features

Although composed of four RRMs, PTB has some noteworthy features. First, the traditional definition of an RRM includes the presence of two ribonucleoprotein (RNP) sequences. These RNP sequences, conserved throughout the protein family, are located centrally on the β-sheet and contain a number of aromatic amino acids generally thought to be important for stacking with RNA bases upon binding. In addition, most RRMs also have a highly conserved glycine at the end of β-strand three. In PTB, however, all four RRMs deviate significantly from this sequence, lacking the aromatic side chains in the RNPs as well as the aformentioned glycine. Yet, all four RRMs have been shown to bind RNA (Oberstrass, et al., 2005). This suggests that the binding mechanism of these RRMs is unique to the family. Understanding the RNA selection and binding mechanism is requisite information for understanding how this very versatile protein functions.

Protein structure

The structures of all four RRMs of PTB have been solved (Conte et al., 2000; Simpson et al., 2004) both alone, and bound to RNA (Oberstrass et al., 2005; Vitali et al., 2006). The protein consists of four RRMs connected by linkers, but also contains some interesting features. RRM1 and RRM4 have the canonical RRM three-dimensional structure, with a four stranded β-sheet packed against two α-helices. However, RRM2 and RRM3 have a C-terminal extension: a β-fifth strand that packs against β-strand two

and is attached to β-strand four with a long, flexible linker that extends across the β-sheet surface. This structural addition not only extends the putative RNA binding surface, but also occludes it, again suggesting that the RNA binding mode for these RRMs is novel.

PTB also exhibits a unique protein-protein interaction between RRM3 and RRM4. While RRM1 and RRM2 appear to exist as structurally individual units, RRM3 and RRM4 interact extensively via a large hydrophobic interface involving both helices of RRM3, large sections of the interdomain linker, and one helix as well as the fourth β-strand of RRM4 (Oberstrass et al., 2005; Vitali et al., 2006). This interaction is present in both the free and RNA bound states, and although the RRMs can be produced as individual units, the interaction does not form when the two individual domains are mixed.

It is common for RRMs to occur in proteins as tandem pairs, although most RRMs exist as structurally independent motifs. Some RRMs, such as sex-lethal (Handa et al., 1999), nucleolin (Allain et al., 2000) and poly-A binding protein, PABP (Deo et al., 1999) are known to associate with other RRMs in the RNA bound state, but not in the unbound state. However, the interaction between PTB RRM3 and RRM4 is unique in terms of its extensive interface, the orientation of the two domains, and its persistence in both free and bound forms.

PTB-RNA interactions

Determination of a binding mechanism for PTB remains elusive, yet much is known about PTB-RNA interactions. The situation is complicated by the fact that this protein is involved in so many different cell processes, and thus, many different RNAs. A good deal of the work that has been done has used one of the many natural targets of PTB, the $GABA_A$ γ2 pre-mRNA, as well as synthetic oligomers. As early as 1997, Perez et al. determined that PTB binds to UCUU motifs, but specificity was observed only in the context of longer pyrimidine tracts. More recent structural studies using tetramer to hexamer polypyrimidines reveal that these short RNAs bind in multiple registers to the RRMs (Oberstrass et al., 2005) underscoring the idea that longer RNA constructs are necessary to obtain consistent binding information.

We now know that the system is even more complicated. Extensive binding studies have shown that the RRMs of PTB have different contributions to binding specificity and affinity (Perez et al., 1997; Clerte and Hall, 2006). Furthermore, RRMs appear to have distinct RNA structural preferences (Clerte and Hall 2009). Of particular interest to the work described here, PTB1:34 prefers unstructured to structured RNA. In addition, Oberstrass et al. (2005) discovered that PTB1:34 binding is optimized when polypyrimidine tracts of the target RNA are separated by 15 nucleotides. Binding stoichiometries also vary widely depending on the RRMs present as well as the RNA used for the binding studies (Clerte and Hall, 2006).

What can be learned from this system?

The fact that this system is made up of well characterized domains, yet has unique features provides an opportunity to ask some fundamental as well as more applied questions pertaining to PTB RNA binding mechanisms; consequences of protein-protein interactions in terms of binding specificity, thermodynamics, and inherent protein properties; and communication between binding sites. In addition, we can exploit the C-terminal extensions, unique RNP sequences and the protein-protein interaction between RRM3 and RRM4 to explore, and hopefully extend, the traditional definition of an RRM.

The ability to address these questions requires engineering and production of appropriate protein constructs as well as thorough biochemical characterization of the constructs. This preliminary work is described in Chapter two. Several NMR techniques are used for more detailed description of the protein properties, including some that are not commonly used. Accurate interpretation of these results requires some background information of the specific experiments used. This background is provided in Chapter three. As will be seen, the way a molecule tumbles in solution is an important component of data analysis and interpretation. Because of this, several methods were used to determine tumbling times of the protein constructs, and these experiments are described in Chapter four. The interaction between PTB RRM3 and RRM4 induces microsecond motions throughout both domains and is required for efficient RNA binding. This work is presented in Chapter five.

A protein construct was prepared in order to address the functional importance of

the RRM3 extension. This protein does not bind RNA, although the structure is not significantly perturbed. These experiments are described in Chapter six. Chemical denaturation experiments revealed that RRM4 has a linear unfolding curve, a property that can be exploited learn more about protein unfolding processes. This protein also turns out to be the largest and first naturally occurring protein to exhibit downhill folding characteristics. These projects are described in Chapter seven. Finally, a mutant predicted to decouple RRM3 and RRM4 without physically separating the domains was engineered. This mutant will be useful for learning more about how the interaction contributes to RNA binding, the nature of the RNA binding site (in the context of PTB1:34 should we think about the RNA binding site as a single site, or two binding sites), as well as addressing questions of allostery and communication paths. The preliminary work done for this project is discussed in Chapter eight.

References

Allain, F. H., Bouvet, P., Dieckmann, T. and Feigon, J. Molecular basis of sequence specific recognition of pre-ribosomal RNA by nucleolin. (2000) *EMBO J.*, **19**, 6870-6881.

Back, S. H., Kim, Y. K., Kim, W. J., Cho, S., Oh, H. R., Kim, J. E. and Jang, S. K. Translation of polioviral mRNA is inhibited by cleavage of polypyrimidine tract-binding proteins executed by polioviral 3Cpro. (2002) *J. Virol.*, **76**, 2529-2542.

Bushell, M., Stoneley, M., Kong, Y.-W., Hamilton, T., Spriggs, K. A., Dobbyn, H. C. and Willis, A. E. Polypyrmidine tract binding protein regulates IRES-mediated gene expression during apoptosis. (2006) *Mol. Cell*, **23**, 401-412.

Castelo-Branco, P., Fugrer, A., Wollerton, M. C., Smith, C. W. J., Moreira, A. and Proudfoot, N. J. Polypyrimidine tract binding protein modulates efficiency of polyadenylation. (2004) *Mol Cell. Biol.* **24**, 4174-4183.

Clerte, C. and Hall, K. B. Characterization of multimeric complexes formed by the human PTB1 protein on RNA. (2006) *RNA*, **12**, 457-475.

Clerte, C. and Hall, K. B. The domains of polypyrimidine tract binding protein have distinct RNA structural preferences. (2009) *Biochemistry*, **48**, 2063-2074.

Coles, L. C., Bartley, M. A., Bert, A., Hunter, J., Polyak, S., Diamond, P., Vadas, M. and Goodall, G. J. A multi-protein complex containing cold shock domain (Y-box) and polylyrimidine tract binding proteins forms on the VEGF mRNA. (2004) *Eur. J. Biochem.*, **271**, 648-660.

Conte, M. R., Grune, T., Kelly, G., Ladas, A., Matthews, S. and Curry, S. Structure of tandem RNA recognition motifs from polypyrimidine tract binding protein reveals

novel features of the RRM fold. (2000) *EMBO J.*, **19**, 3132-3133.

Cornelis, S., Tinton, S. A., Schepens, B., Bruynooghe, Y. and Beyaert, R. UNR translation can be driven by an IRES element that is negatively regulated by polypyrimidine tract binding protein. (2005) *Nucleic Acids Res.*, **33**, 3095-3108.

Cote, C. A., Gatreau, D., Denegre, J. M., Kress, T. L., Terry, N. A. and Mowry, K. L. A *xenopus* protein related to hnRNP1 has a role in cytoplasmic RNA localization. (1999) *Mol. Cell*, **4**, 431-437.

Deo, R. C., Bonanno, J. B., Sonenberg, N. and Burley, S. K. Recognition of polyadenylate RNA by the poly(A)-binding protein. (1999) *Cell*, **98**, 835-845.

Ghetti, A., Pinol-Roma, S., Michael, W. M., Morandi, C. and Dreyfuss, G. hnRNP1, the polypyrimidine tract binding protein: distinct nuclear localization and association with hnRNAs. (1992) *Nucleic Acids Res.*, **20**, 3671-3678.

Gontarek, R. R., Gutshall, L. L., Herold, K. M., Tsai, J., Sathe, G. M., Mao, J., Prescott, C and Del Vecchio, A. M. hnRNPC and polypyrimidine tract binding protein specifically interact with the pyrimidine rich region within the 3'NTR of the HCV RNA genome. (1999) *Nucleic Acids Res.*, **27**, 1457-1463.

Gosert, R., Chang, K. H., Rijnbrand, R., Yi, M. Y., Sangar, D. V. and Lemon, S. M. Transient expression of cellular polypyrimidine tract binding protein stimulates cap-independent translation directed by both picornaviral and flaviviral internal ribosome entry sites *in vivo*. (2000) *Mol. Cell. Biol.*, **20**, 1583-1595.

Handa, N., Nureki, O., Kurimoto, K., Kim, I., Sakamoto, H., Shimura, Y., Muto Y. and Yokoyama, S. Structural basis for recognition of the tra mRNA precursor by the Sex-lethal protein. (1999) *Nature*, **398**, 579-585.

Hamilton, B. J., Genin, A., Cron, R. Q. and Rigby, W. F. C. Delineation of a novel pathway that regulates CD154 (CD40 ligand) expression. (2003) *Mol. Cell. Biol.*, **23**, 510-525.

Hellen, C. U. T., Witherell, G. W., Schmid, M., Shin, S. H., Pestova, T. V., Gil, A. and Wimmer, E. A cytoplasmic 57-kDa protein that is required for translation of picornavirus RNA by interal ribosomal entry is identical to the nuclear pyrimidine tract binding protein. (1993) *Proc. Natl. Acad. Sci.*, **90**, 7642-7646.

Kanda, T., Gauss-Muller, V., Cordes, S., Tamura, R., Okitsu, K., Shuang, W., Nakamoto, S., Fujiwara, K., Imazeki, F and Yokosuka, O. Hepatitis A virus (HAV) proteinase 3C inhibits HAV IRES-dependent translation and cleaves the polpyrimidine tract binding protein. (2009) *J. Virol. Hep.*, xx,xxxx.

Ma, S., Liu, G., Sun, Y. and Xie, J. Relocalization of the polypyrimidine tract binding protein during PKA induced neurite growth. (2007) *Biochim. Biophys. Acta.*, **1773**, 912-923.

Mitchell, S. A., Spriggs, K. A., Bushell, M., Evans, J. R., Stoneley, M., le Quesne, J. P. C., Spriggs, R. V. and Willis, A. E. Identification of a motif that mediates polypyrimidine tract binding protein dependent internal ribosome entry. (2005) *Genes Dev.*, **19**, 1556-1571.

Li, B. and Yen, T. S. B. Characterizatin of the nuclear export signal of polypyrimidine tract binding protein. (2002) *J. Biol. Chem.*, **277**, 10306-10314.

Lou H., Helfman, D. M., Gagel, R. F. and Berget S. M. Polypyrimidine tract binding protein positively regulates inclusion of a alternative 3'-terminal exon. (1999) *Mol. Cell. Biol.*, **19**, 78-85.

LeSommer, C., Lesimple, M., Mereau, A. S., Menoret, S., Allo, M. R. and Hardy, S. PTB regulates the processing of a 3'-terminal exon by repressing both splicing and polyadenylation. (2005) *Mol. Cell. Biol.*, **25**, 9595-9607.

Oberstrass, F. C., Auweter, S. D., Erat, M., Hargous, Y., Henning, A., Wenter, P., Reymond, L., Amir-Ahmady, B., Pitsch, S., Black, D. L. and Allain, F. H.-T. Structure of PTB bound to RNA: Specific binding and implications for splicing regulation. (2005) *Science*, **309**, 2054-2057.

Pautz, A., Linker, K., Hubrich, T., Korhonen, R., Altenhofer, S. and Kleinert, H. The polypyrimidine tract binding protein (PTB) is involved in the post-transcriptional regulation of human inducible nitric oxide synthase expression. (2006) *J. Biol. Chem.*, **81**, 32294-32302.

Perez, I., McAfree, J. G. and Patton, J. G. Multiple RRMs contribute to RNA binding specificity and affinity for polypyrimidine tract binding protein. (1997) Biochemistry, **36**, 11881-11890.

Perez, I., Lin, C-H., McAfee, J. G. and Patton, J. G. Mutation of PTB binding sites causes misregulation of alternative 3' splice site selection in vivo. (1997) *RNA*, **3**, 764-778.

Pilipenko, E. V., Pestova, T. V., Kolupaeva, V. G., Khitrina, E. V., Poperechnaya, A. N., Agol, V. I. and Hellen, C. U. T. A cell cycle dependent protein serves as a template-specific translation initiation factor. (2000) *Genes Dev.*, **14**, 2028-2045.

Shen, H. H., Kan, J. L. C., Ghigna, C., Biamonti, G. and Green, M. R. A single polypyrimidine tranct binding protein (PTB) binding site mediates splicing inhibition at mouse IgM exons M1 and M2. (2004) *RNA*, **10**, 787-794.

Simpson, P. J., Monie, T. P., Szendroi, A., Davydova, N., Tyzack, J. K., Conte, M. R., Read, C. M., Cary, P. D., Svergun, D. I., Konarev, P. V., Curry, S. and Matthews, S. Structure and RNA interactions of the N-terminal RRM domains of PTB. (2004) *Structure*, **12**, 1631-1643.

Song, Y. T., Tzima, E., Ochs, K., Bassili, G., Trusheim, H, Linder, M., Preissner, K. T. and Niepmann, M. Evidence for an RNA chaperone function of polypyrimidine tract binding protein in picornaviruns translation. (2005) *RNA*, **11**, 1809-1824.

Tillmar, L., Carlsson, C. and Welsh, N. Control of insulin mRNA stability in rat pancreatic islets: regulatory role of a 3'-untranslated region pyrimidine rich sequence. (2002) *J. Biol. Chem.*, **277**, 1099-1106.

Valcarcel, J. and Gebauer, F. Post-transcriptional regulation: the dawn of PTB. (1997) *Curr. Biol.* **7**, 705-708.

Vitali, F., Henning A., Oberstrass, F. C., Hargous, Y., Auweter, S. D., Erat, M. and Allain, F. H.-T. Structure of the two most C-terminal RNA recognition of PTB using segmental isotope labeling. (2006) *EMBO J.*, **25**, 150-162.

Wollerton, M. C., Gooding, C., Wagner, E. J., Garcia-Blanco, M. A. and Smith, C. W. J. Autoregulation of polypyrimidine tract binding protein by alternative splicing leads to nonsense mediated decay. (2004) *Mol Cell*, **13**, 91-100.

Xu, M. G. and Hecht, N. B. Polypyrimidine tract binding protein 2 stabilizes phosphoglycerate kinase 2 mRNA in murine male germ cells by binding to its 3'-UTR. (2007) *Biol. Reprod.*, **76**, 1025-1033.

Yuan, X., Davydova, N., Conte, M. R., Curry, S. and Matthews, S. Chemical shift mapping of RNA interactions with the polypyrimidine tract binding protein.

(2002) *Nucleic Acids Res.*, **30**, 456-462.

Zang, W. Q., Li. B., Huang, P. Y., Lai, M. M. C. and Yen, T. S. B. Role of polypyrimidine tract binding protein in the function of the hepatitis B virus posttranscriptional regulatory element. (2001). *J. Virol.*, **75**, 10779-10786.

Chapter 2

Design, production, and biochemical characterization of protein constructs.

Investigation of the functional importance of the unique interaction between the two C-terminal RRMs of PTB depends on the ability to study the RRMs both individually and in concert. Disruption of the hydrophobic core between the two domains required mutation of 6 of the 27 amino acids involved in the interface, replacing hydrophobic with charged side chains: I33K, F123E, I126K, E179K, V182E, and I186K (Vitali et al., 2006). To avoid such drastic mutagenesis, which may impact the nature of the protein structure, properties and function in addition to disruption of the interface, the individual RRMs were prepared as separate entities for this work.

Protein constructs were prepared to facilitate comparison of the properties of the individual domains, RRM3 and RRM4, to the interacting domains, PTB1:34. Truncations of PTB1:34 were designed as follows. Numbering of residues uses the numbering scheme adopted in PDB ID:2EVZ throughout (Figure 1). The entire PTB1:34 sequence contains 198 amino acids, numbered 10-208. RRM3 consists of residues 10-116, and RRM4 is made up of residues 114-208. The absence of three residues when the individual domains are mixed is not ideal, however, two other constructs truncated in a manner to avoid such a gap did not express in the system used.

Truncated protein constructs were prepared by subcloning of the PTB1:34 sequence. RRM3 was designed and cloned prior to my arrival in the lab. Initial attempts

to obtain RRM4 by cutting before G108 and F115 were unsuccessful. A third attempt to truncate the protein prior to S120 was successful, using top primer 5'-CATGCCATGGGATCCAAGAACTTCCAGAACATATTCCCG-3' and PTB1:34 bottom primer 5'- (As far as I can tell, Pat Jett originally designed this primer, but I can find no mention of this primer in her records.) Plasmids were transformed into BL21-DE3 gold *E. Coli* for protein expression induced with 1mM IPTG at optical density of 0.75, and grown for 4 hours at 37°C in appropriate media. Unlabeled proteins were grown in LB, and labeled proteins were grown in M9 minimal media using ^{15}N-ammonium chloride and ^{13}C-glucose as the sole nitrogen and carbon sources as needed. Cells were harvested by centrifugation at 6500 rpm in a Sorvall GSA rotor (6876 g), washed in GI buffer (20 mM Tris, pH 7.5, 20 mM NaCl and 2 mM EDTA), repelleted and stored at -80°C overnight. Pellets were thawed and re-suspended in 25 mL of buffer B (20 mM NaOAc, pH 5.3, 200 mM NaCl, with the exception of RRM4, where 50 mM NaCl was used instead, and 2 mM EDTA), with 120 μL of 20 mg/mL PMSF in isopropanol, 1mL 10X Sigma protease inhibitor cocktail and 250 μL 2U/μL DNase II. Cell lysate was centrifuged at 15,000 rpm in a Sorvall SS34 rotor for 20 minutes and the supernatant was dialyzed against 1L of buffer B for 2-4 hours at room temperature. Dialysis product was cleared by centrifugation and purified using a single CM-sepharose ion exchange column with a NaCl gradient (100-500 mM for PTB1:34, 50-100 mM for RRM3 and 10-500 mM for RRM4). Proteins were concentrated using Vivaspin centrifugal concentrators (10K MWCO for PTB1:34, 5K MWCO for RRM3 and 3K MWCO for RRM4), and purity assessed using SDS-PAGE. Protein concentration was quantified by absorption at 280 nm for PTB1:34 and RRM3 (Extinction coefficient of 2560 M^{-1} cm^{-1}), and absorption at

250 nm for RRM4 (Extinction coefficient of 1200 M^{-1} cm^{-1}).

Secondary structure analysis of the protein constructs

Circular dichroism was used to monitor secondary structure as an initial check that the prepared protein constructs were folded. Spectra consistent with folded proteins were obtained for all constructs, although the magnitude of the signal for PTB1:34, adjusted for number of residues, was less than expected (Figure 2). Importantly, the signal for mixed RRM3 and RRM4 was much greater than that for PTB1:34, and consistent with the signal for the individual domains alone. The CD data are presented in units of molar ellipticity per residue. Data from the polarimeter are given as ellipticity, θ, in mdeg cm^2 per dmol. The molar ellipticity, θ_R, which takes into account protein concentration as well as number of residues is:

$$\theta_R = (100\ \theta) / (C\ l\ n)$$

where C is the molar protein concentration, n is the number of residues, and l is the path length in cm. The factor of 100 is included for historical reasons.

The discrepancy between the apparent amount of secondary structure in PTB1:34 and the mixed individual domains suggests a fundamental change in the inherent protein properties that takes place when RRM3 and RRM4 interact, and that this interaction does not occur when the individual domains are mixed. These data imply that transient secondary structure formation, reminiscent of a molten globule, is a property of PTB1:34 that is lost upon separation of the domains, with no accompanying change in overall fold.

In an attempt to compare the thermodynamics of the protein constructs, the secondary structure was monitored using a spectropolarimeter at 222 nm during thermal unfolding. Unfortunately, PTB1:34 and RRM3 melts were irreversible, and RRM4 did not melt, even at temperatures greater than 90°C. Constrained by the system, chemical denaturation was employed instead. Urea was not an effective denaturant and PTB1:34 remained folded even at 7M Urea. Guandine HCl thus became the denaturant of choice since it was able to denature all protein constructs (Figure 2), although the curves generated were not amenable to quantitative comparison. While RRM3 denaturation generated a curve that could easily be fit by a 2-state model, RRM4 and PTB1:34 did not. The curve for PTB1:34 had non-linear baselines in different directions, and RRM4 was mostly linear throughout. The linear denaturation profile of RRM4 is discussed in detail in Chapter 7.

Despite the inability to obtain quantitative thermodynamic information, these data bring to light interesting properties of the protein constructs. Namely, the multi-state unfolding curve for PTB1:34 could be due to an initial separation of the domains followed by unfolding of the individual domains. Furthermore, the linear denaturation curve of RRM4 implies a series of incremental energy barriers along the unfolding pathway, suggestive of a protein that could be described as a downhill folder. Gdn HCl denaturation was, for the most part, reversible, with most of the signal for all protein constructs recovered after dialyzing 6 M Gdn against 0 M Gdn (1:10,000) overnight (Figure 3).

Comparison of RNA binding efficiency

Electrophoretic mobility shift assay was used to determine the relative binding efficiencies of the interacting versus individual RRMs. The RNA used for these studies is from the GABA intron mRNA and contains two polypyrmidine tracts separated by a series of 15 adenosines (Figure 4). This RNA is predicted by m-fold (Zucker, 2003) to be mostly unstructured, and the 15 nucleotide spacer between the polypyrimidine tracts has been shown to be optimal for RNA binding (Oberstrass, et al. 2005). Previous work in the lab has shown that the stoichiometry of binding this RNA to PTB1:34 is 1:8 (Clerte and Hall, 2006) precluding isothermal titration calorimetry as an alternative method for measuring binding affinities. Although quantitative comparisons are not possible, it is useful to consider the possible outcomes in terms of an equilibrium association constant, K_a. For a single binding site, the binding reaction can be written as:

$$P + L \leftrightarrow PL$$

Where P indicates the protein, and L the ligand, in this case the RNA, with an equilibrium association constant K_a. Comparison of single site binding reactions is straightforward. However, for a macromolecule with two completely independent and identical binding sites:

$$P + 2L \leftrightarrow PL_2$$

Mass action dictates that the overall binding association constant, K_1, contains the product of the equilibrium constants for each site:

$$K_1 = K_a K_b$$

where K_a is the binding association constant for the first site, and K_b is the binding association constant for the second site. Thus, if both sites bind a ligand, even if one of

the sites binds with a very low affinity, physical separation of the domains, so that each molecule has only a single binding site, may significantly reduce the apparent binding association constant without actually lowering the binding affinity of either site. In other words, one expects the equilibrium binding association constant for a two site molecule to be significantly larger than the sum of two single site binding constants since the stepwise binding constant contains the product of each independent binding constant.

Alternatively, it is important to consider the possibility that the binding surface of PTB1:34 may be better described as a single binding site, allowing a direct comparison of the individual RRMs versus PTB1:34 binding affinities. Lamichhane et al. (2010) have shown that the PTB1:34 RNA binding site is physically continuous throughout RRM3, RRM4 and the interdomain linker. In addition, the interaction between the two RRMs appears to alter the RNA binding affinity even though none of the residues involved in the RRM3/RRM4 interaction are located in the RNA binding site, suggesting that the interaction between the RRMs is an allosteric effector of protein properties that contributes to binding. It is clear that the interaction between RRM3 and RRM4 influences the RNA binding properties of the molecule although the mechanism of this alteration is not known. Because it is not understood how the interaction influences the properties of the molecule that contribute to RNA binding, we must consider the possibility that the RNA binding surfaces of RRM3 and RRM4 combine in PTB1:34 to form a single, albeit larger, RNA binding site. Thus, the idea that the two RRMs come together to form a single, unique RNA binding site should be considered a viable alternative to the interpretation of binding experiments.

EMSA data for each protein construct as well as a mixture of the individual domains is shown in figure 4. While PTB1:34 binds even at the lowest protein concentration tested, 100 nM, RRM4 does not bind at all, even at the highest protein concentration tested (10μM). RRM3 binds with about a 50-fold lower affinity than PTB1:34, with the first shift observed at around 500 nM. The actual binding affinities cannot be determined using these experiments due to smeared bands and complicated stoichiometries. Nonetheless, consideration of a couple of scenarios is useful for qualitative interpretation of the results. In terms of binding constants, if RRM4 does bind at all, even at very high concentrations, the overall binding constant for two independent binding sites would indicate significantly weaker binding than that observed for PTB1:34, due simply to the separation of the domains without any inherent loss of affinity. In other words, combining two independent binding sites in a single molecule, even if one of them binds very weakly, is expected to result in an entity that binds RNA with apparent greater efficiency proportional to the product of each site's binding association constant. Thus, the fact that the mixture of RRM3 and RRM4 binds similarly to RRM3 alone should not be surprising, but does make clear that the interaction between RRM3 and RRM4 that confers enhanced binding efficiency does not occur simply by mixing the domains. However, we also need to consider the possibility that PTB1:34 is better described as a molecule with a single RNA binding site. In this case, a direct comparison between the binding abilities of the individual RRMs and PTB1:34 can be made, showing a drastic increase in RNA binding efficiency due to the RRM3/RRM4 interaction.

Analysis of surface electrostatic potential

To gain a more detailed understanding of how these charged molecules interact, it is important to consider the potential electrostatic component of PTB1:34 binding to RNA. To this end, the electrostatic potential was calculated and mapped onto the solvent accessible surface area (SASA) of RRM3, RRM4, and PTB1:34 using the Adaptive Poisson Boltzman Solver (APBS) (Baker, et al. 2001) and Pymol (Figure 5). These results show that the overall charge distribution is slightly rearranged when RRM3 and RRM4 interact, which may contribute to RNA binding by providing a positively charged path that wraps around PTB1:34. In addition, a large negative patch on RRM4 remains, and is even possibly concentrated in the context of PTB1:34, suggesting a possible protein-protein interaction site.

Structural characterization of protein constructs

The above data provide evidence that the interaction between the individual domains that comprise PTB1:34 generates changes in protein properties that appear to be important for RNA binding efficiency. However, the denaturation studies make clear that the interaction does not significantly impact the thermodynamic stability or the ability of the RRMs to fold. Taken together, these results raise a few questions. First, if the changes in binding ability cannot be explained in terms of thermodynamic stability, folding propensity, or pure electrostatic potential distribution, what exactly are the properties that change upon interaction of RRM3 and RRM4? Also, does RRM4 contribute to binding in the context of PTB1:34? In addition, the RNA binding sites lie on opposing faces of PTB1:34, distal from the RRM3/RRM4 interface, implying

allostery. What is the mechanism of communication from the interface to the RNA binding site? The first questions were addressed in a detailed, residue specific manner, using ^{15}N-NMR to probe the structural and dynamic properties of each protein construct. The question of allostery will be addressed in chapter 8.

NMR studies of RRM3, RRM4 and PTB1:34 commenced with comparisons of the ^{15}N-HSQC spectra (figure 6). Further exploration requires that most amides in the protein backbones give rise to single, well resolved peaks, and that the spectra have adequate chemical shift dispersion consistent with folded, stable proteins. For the most part, this was found to be the case, although a few more peaks than expected were observed in RRM3, suggesting exchange that is slow on the NMR timescale, resulting in multiple resonances for some of the amides. The degree of chemical shift dispersion was consistent with that expected for folded proteins.

^{1}H/^{15}N-HSQC spectra were overlaid to compare PTB1:34 to the individual domains and reveal differences in many of the chemical shifts. However, because separation of the domains is expected to change the nuclear shielding of amides involved in the extensive interface, the difference in chemical shifts does not necessarily imply structural changes. In order to determine the magnitude of the chemical shift changes, and thus the possibility of significant structural changes, assignment of the peaks in each ^{15}N-HSQC spectrum is necessary.

The structure of PTB1:34 has been solved (Conte, et al., 2000; Oberstrass, et al. 2005), but under different solution conditions than were deemed appropriate for our RNA binding experiments. Nonetheless, attempts to use these previously determined resonance assignments were made by extrapolating chemical shifts from published to our solution conditions. The chemical shifts of PTB1:34 were mapped from pH 5.2 to 6.8, and from salt concentrations from 0-150 mM KCl. Unfortunately, this approach did not map to the assigned resonances in PTB1:34, so we had to repeat the assignment process. The structures of RRM3 and RRM4 have not been solved, so we assigned the amide resonances of those constructs as well. $^{13}C/^{15}N$ labeled proteins were prepared, and standard CBCA(CO)NH/HNCACB and HNCO/HNCACO experiments were run and then processed and analyzed using Felix (Accelyrs). Multiple resonances observed for some loop regions of PTB1:34 and RRM4, and several loop regions of RRM3 did not allow full assignments to be made. Assignments were identified for over 85% of PTB1:34 and RRM4 residues, and over 65% of RRM3.

Once the backbone amides for RRM3, RRM4 and PTB1:34 were assigned, chemical shift differences between the individual and interacting RRMs, $\Delta\delta$ were calculated for each residue in terms of the proton chemical shift:

$$\Delta\delta = [(\delta_{Ha} - \delta_{Hb})^2 + ((\delta_{Na} - \delta_{Nb})/10)^2]^{1/2}$$

Where δ_{Ha} is the proton chemical shift of the individual domain, δ_{Hb} is the proton chemical shift of the interacting domains, δ_{Na} is the chemical shift in the nitrogen dimension for the individual domain and δ_{Nb} is the chemical shift in the nitrogen dimension for the interacting domains. These results were mapped onto the three

dimensional structure of PTB1:34 (PDB ID:2EVZ) (Figure 7) and clearly show that most of the observed chemical shift changes are small and concentrated at the RRM3/RRM4 interface, revealing that there are no major structural rearrangements that occur as a result of the interaction, and the observed differences in chemical shifts are due mainly to changes in the nuclear environment at the interface.

Assignment of the amide resonances in PTB1:34 also allow residue specific RNA binding studies, if the system permits. Stoichiometry experiments for PTB1:34 show that it binds the GABA intron RNA 8:1 (Clerte and Hall, 2006), a system that is not amenable to NMR experiments. Furthermore, Oberstrass, et al., have solved the structure of PTB1:34 bound to RNA, but required RNA no longer than tetramer and low salt buffer, conditions which are not conducive to specific RNA binding, to obtain adequate NMR spectra. We designed an RNA that mimicked a natural target of PTB1:34 for NMR binding studies using more appropriate conditions. This RNA consists of 11 nucleotides (UUCUCUUUUCU) which we hoped to be long enough to span the entire RNA binding surface of PTB1:34, but still bind specifically and with a 1:1 stoichiometry. Previous EMSAs using a truncated (38nt) version of the GABA intron RNA did not yield shifts at protein concentrations up to 10 μM, so this RNA is not expected to bind under the same conditions. However, we did expect this RNA to bind at the much higher concentrations used for NMR.

A fresh sample of ^{15}N-PTB1:34 was prepared and titrated with the designed RNA from 0.5:1 to 2:1 (Figure 8). In the presence of RNA, several resonances are shifted, but

more resonances disappear, suggesting that these resonances go into exchange due either to the RNA coming on and off the protein, or by the creation of a dynamic binding site (the protein loses rigidity as it binds to flexible RNA). Increasing the RNA concentration did not cause the invisible resonances to reappear, meaning that, if the observed exchange is due to transient RNA binding, saturating the protein with RNA does not drive the binding equilibrium to a completely bound state. Future experiments of this sort may be more informative with an RNA that contains two short polypyrimidine tracts separated by 15 adenosines.

Qualitatively, the results are consistent with the published structure of PTB1:34 bound to RNA (Oberstrass et al, 2005). That is, the same amides were found to be involved in RNA binding, despite the different ligands and solution conditions. Importantly, residues throughout the protein including sections of RRM3, RRM4 and the interdomain linker are affected by the addition of RNA. Clearly, RRM4 does contribute to the binding of PTB1:34 to this RNA.

The above structural characterization confirms that all protein constructs are folded and thermodynamically stable, and that RRM4 does contribute to binding in the context of PTB1:34. Yet, it remains to be determined exactly what inherent property of the system changes as a result of the RRM3/RRM4 interaction, and how this information is propagated from the RRM3/RRM4 interface to the RNA binding sites. The fact that PTB1:34 binds preferentially to flexible RNA (Clerte and Hall, 2009), and that several resonances go into exchange with the addition of RNA, accompanied by the observation

of multiple resonances in the RRM3 ^{15}N-HSQC spectrum, lead to the speculation that protein dynamics play an important role in the function of this molecule. NMR relaxation experiments provide access to residue specific motions on timescales from picoseconds to seconds, and is thus the method of choice for further investigation of this system.

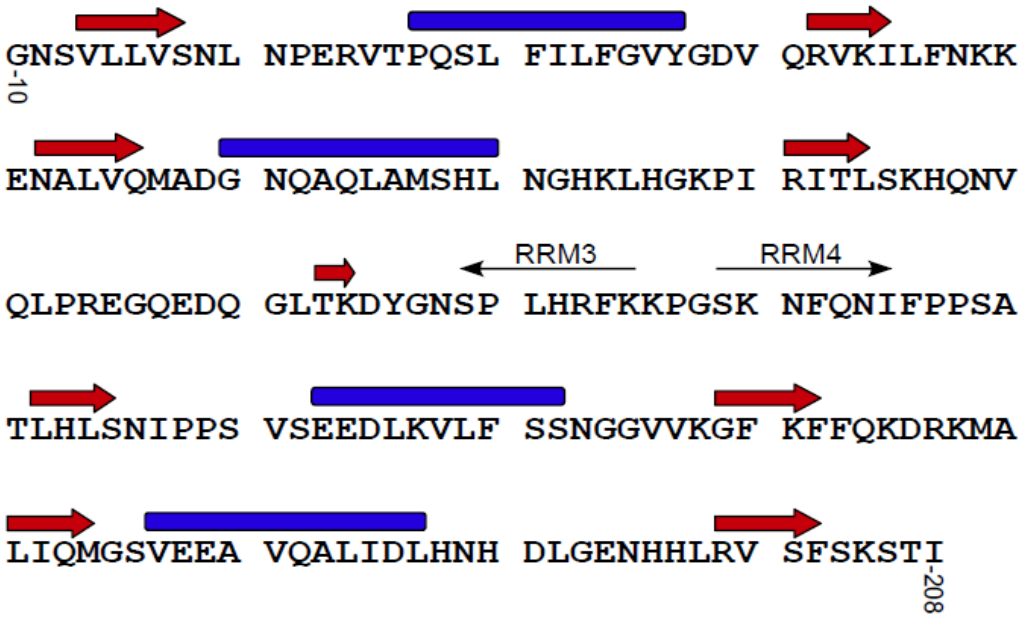


Figure 1: Sequence of PTB1:34, RRM3 and RRM4. Amino acid sequence using the numbering scheme adopted in PDB ID:2EVZ. Secondary structure is depicted above the sequence with beta strands as red arrows and alpha helices as blue cylinders. Individual domains RRM3 and RRM4 were prepared by truncation as indicated by the black arrows.

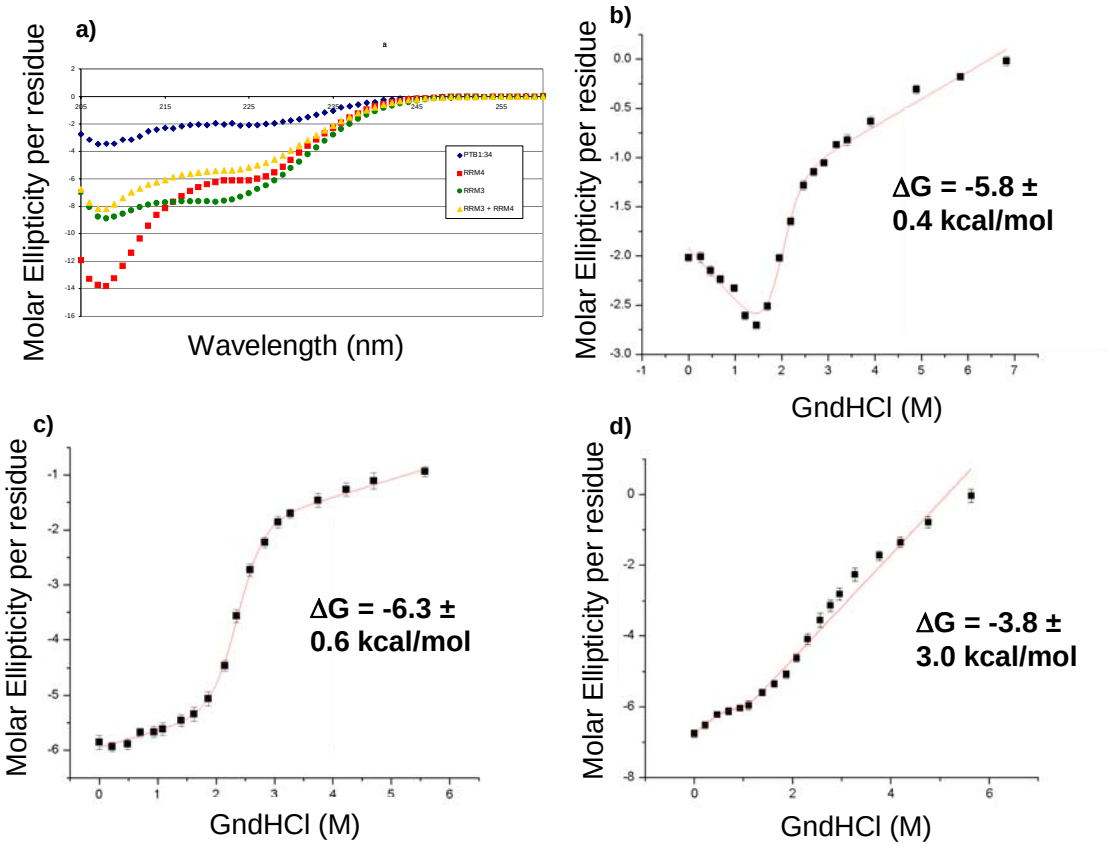


Figure 2: All protein constructs used are folded, yet have significantly different CD spectra and unfolding profiles. Circular dichroism (CD) spectra of each protein construct, as well as an equimolar mixture of RRM3 and RRM4 are shown in (**a**), with PTB1:34 in blue, RRM3 in green, RRM4 in red, and the RRM3/RRM4 mixture in yellow, indicating that all species are folded, and confirms that mixing the individual RRMs does not form PTB1:34. GndHCl denaturations were monitored by CD at 222 nm, and fit with a 2-state unfolding model (red lines) for PTB1:34 (**b**), RRM3 (**c**), and RRM4 (**d**). While RRM3 can be fit with this model, RRM4 and PTB1:34 cannot, suggesting that the thermodynamic properties of the constructs change as a consequence of the RRM3/RRM4 interaction.

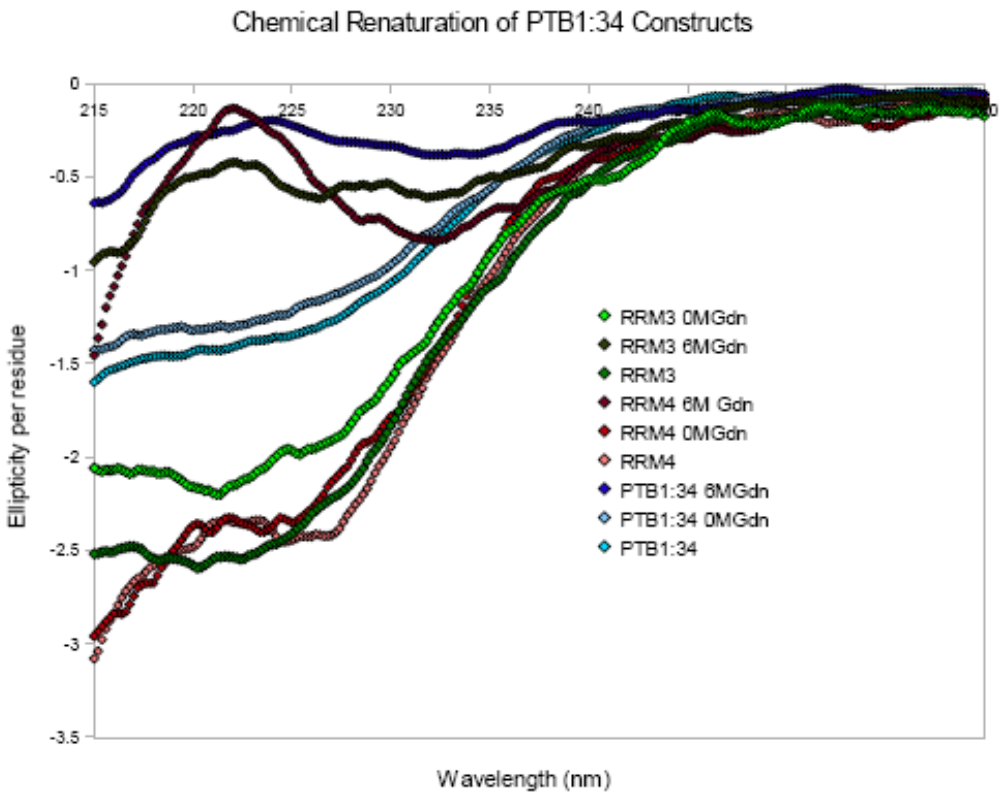


Figure 3: Renaturation of chemically denatured protein constructs. Comparison of CD spectra for native, denatured and renatured RRM3 (green), RRM4 (red) and PTB1:34 (blue). Native proteins were denatured by equilibration in 6 M GndHCl overnight at room temperature. The same protein samples were dialyzed against 1:10,000 buffer with 0 M GndHCl overnight for renaturation. While RRM4 regained all of its original CD signal, PTB1:34 and RRM3 did not. However, the losses in signal observed for RRM3 and PTB1:34 are consistent with the decreased signal observed over time typically experienced when working with these proteins.

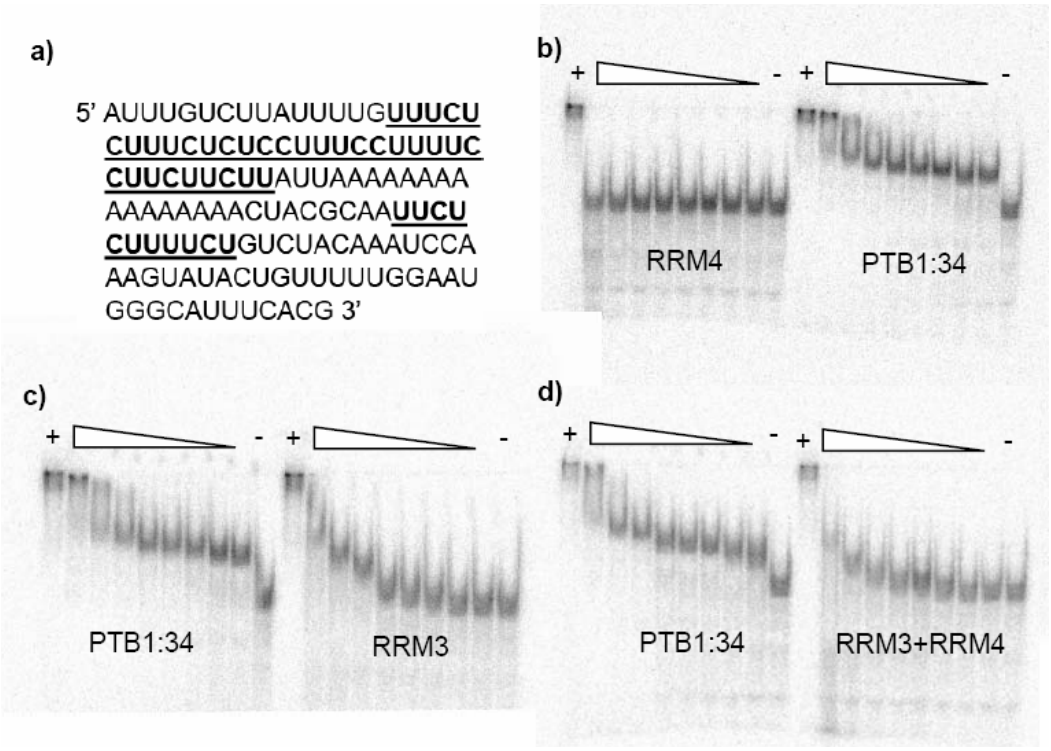


Figure 4: The binding affinity of PTB1:34 is more than the sum of its parts.

Electrophoretic mobility shift assays were used to compare the relative binding affinities of the $GABA_A$ γ2 pre-mRNA intron (**a**) to the PTB1:34 protein constructs. PTB1:34 binds at the lowest protein concentration tested, 10 nM, while RRM4 does not bind at all, even at the highest concentration tested, 5 μM (**b**). A similar comparison in (**c**) shows that RRM3 does bind to this RNA, but with around 50-fold lower affinity than PTB1:34, as the first significant band shift does not occur at protein concentrations less than 500 nM. Mixing RRM3 and RRM4 does not rescue the RNA binding (**d**), since an equimolar mixture of the two domains binds with affinity similar to that of RRM3 alone. All EMSAs were run at 4° C, and included a lane with RNA only as a negative control, and a lane with 800 nM full-length PTB, which is known to bind to this RNA with high affinity, as a positive control.

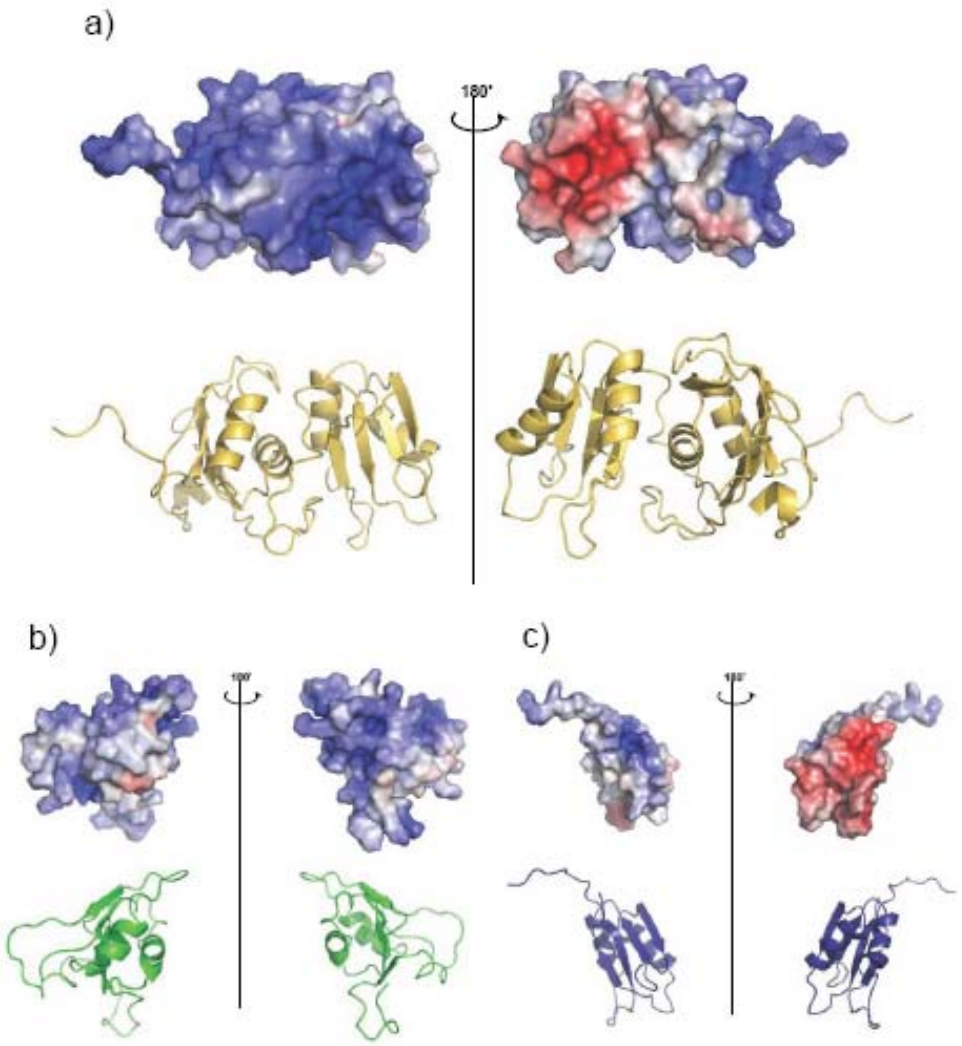


Figure 5: Electrostatic potentials of PTB1:34 and the individual domains may contribute to their functional differences. Electrostatic potential mapped onto the solvent accessible surface area of PTB1:34 (**a**), RRM3 (**b**), and RRM4 (**c**), with positive patches shown in blue, and negative patches in red, shows that the interaction between RRM3 and RRM4 organizes the charge distribution of the protein, and may be important to protein function.

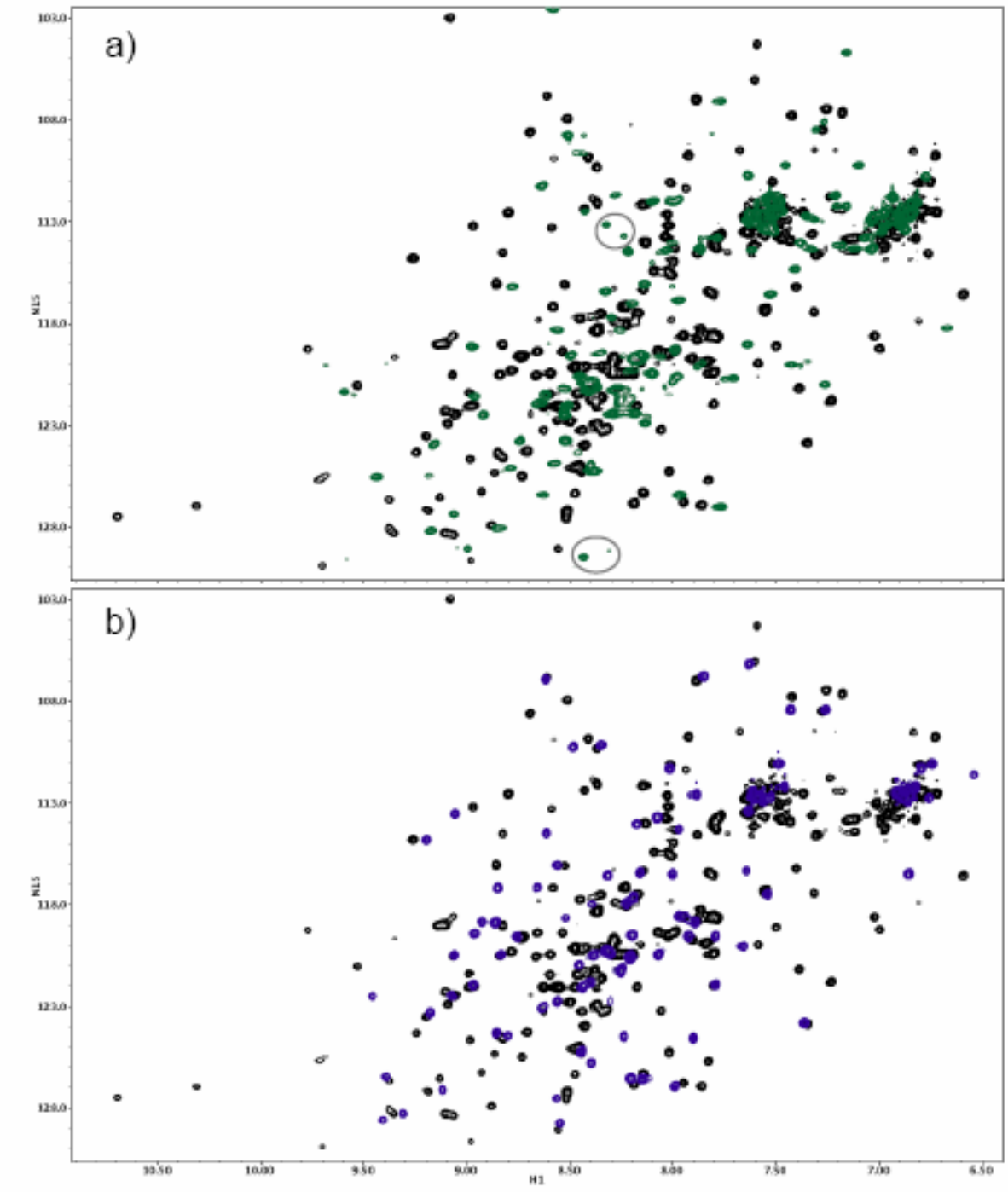

a)
b)
103.0
108.0
113.0
118.0
123.0
128.0
N15
10.50
10.00
9.50
9.00
8.50
8.00
7.50
7.00
6.50
H1

Figure 6: $^{1}H/^{15}N$-HSQC spectra comparisons show that all protein constructs are folded, but have different structural properties. A comparison of RRM3 and PTB1:34 (**a**), shows that almost none of the RRM3 peaks (green) are superimposable with the peaks from PTB1:34 (black), indicating global differences in environment, though not necessarily structure. In addition, many of the RRM3 peaks are broadened, and several residues have multiple resonances (circles), indicating structural heterogeneity on the chemical shift time scale. Conversely, RRM4 (blue) has a single resonance for every amide (**b**), and the peaks are uniformly narrow. Most of the RRM4 peaks are superimposable onto the spectrum of PTB1:34, though many are shifted. PTB1:34 displays neither the narrow line-widths of RRM4 nor the multiple resonances of RRM3, but rather behaves as a unique entity.

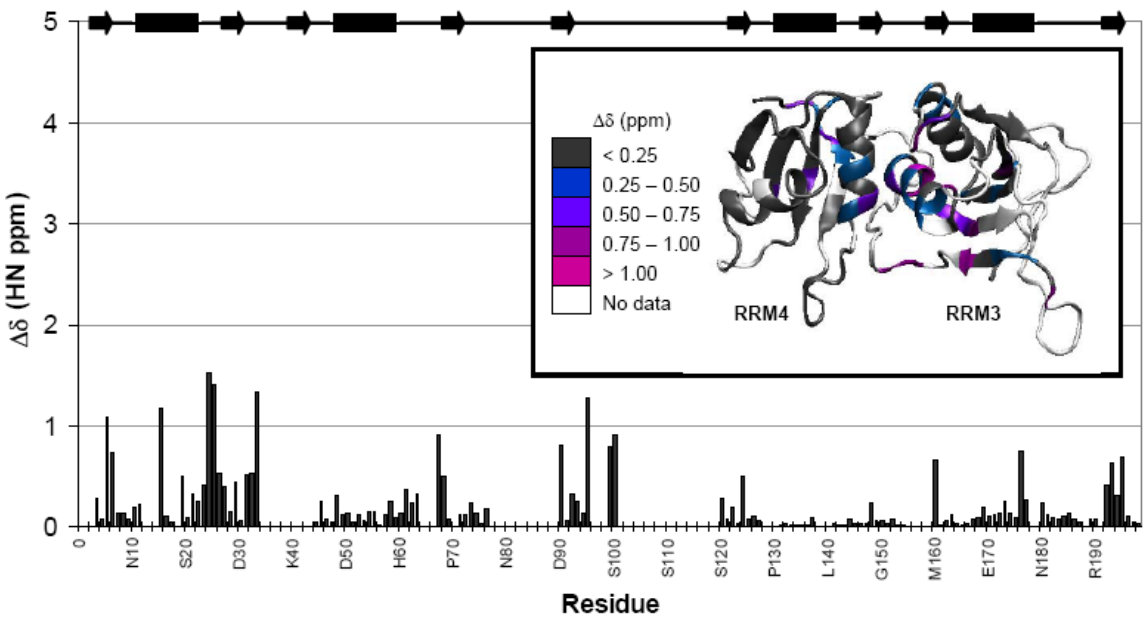


Figure 7: Changes in chemical shifts between the individual and interacting RRMs indicate only minor structural changes when the domains are separated. The change in chemical shift, Δδ, in terms of proton ppm, show that the majority of differences are concentrated at the RRM3/RRM4 interface. Δδ is shown as bars with the protein secondary structure indicated by bars (α-helix) and arrows (β-strand) along the top of the plot. For visual clarity, these changes are mapped onto the structure of PTB1:34 (inset), where white shows the areas of the protein were no data were available, grey indicates no significant Δδ (< 0.25 ppm), blue indicates Δδ between 0.25 and 0.50 ppm, violet, Δδ between 0.50 and 0.75 ppm, purple between 0.75 and 1.00 ppm, and magenta shows the most significant Δδ of greater than 1.00 ppm. Many residues in the interdomain linker are expected to have significant chemical shift changes due to altered environment but could not be calculated since the linker region was largely unassignable for the individual domains.

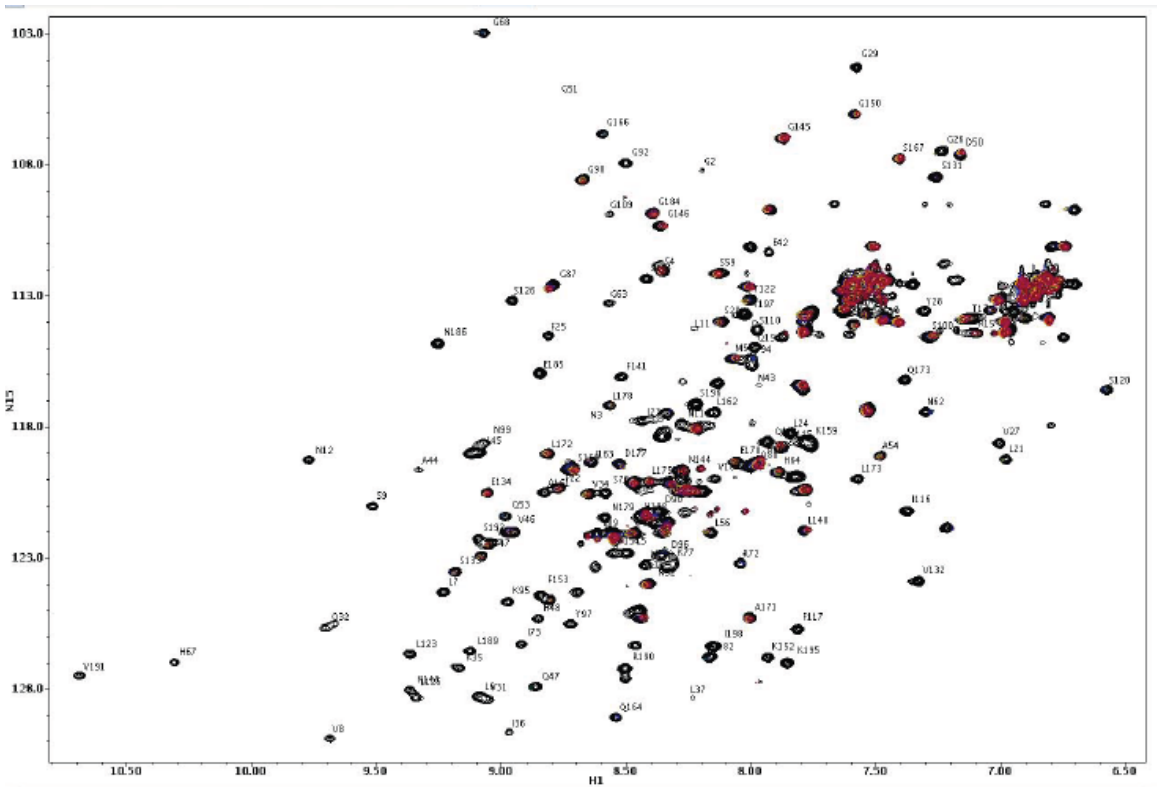


Figure 8: Binding of 11mer RNA to PTB1:34 involves residues in RRM3, RRM4 and the interdomain linker. Overlay of $^{1}H/^{15}N$-HSQC spectra collected for protein alone (black), and with increasing amounts of RNA. Molar ratios of RNA to:protein are shown in blue (0.5:1), green (1:1) and red (2:1). Most residues that are affected by the binding disappear rather than shift, making characterization of the bound PTB1:34 impossible. No significant changes are observed as the amount of RNA is increased, suggesting that even at the lowest amount of RNA used (0.5:1) binding sites in both RRM3 and RRM4 are saturated.

References:

Baker, N. A., Sept, D., Joseph, S., Holst, M. J. and McCammon, J. A. Electrostatics of nanosystems: application to microtubules and the ribosome. (2001) *Proc. Natl. Acad. Sci.*, **98**, 10037-10041.

Clerte, C. and Hall, K. B. Characterization of multimeric complexes formed by the human PTB1 protein on RNA. (2006) *RNA*, **12**, 457-475.

Clerte, C. and Hall, K. B. The domains of polypyrimidine tract binding protein have distinct RNA structural preferences. (2009) *Biochemistry*, **48**, 2063-2074.

Conte, M. R., Grune, T., Kelly, G., Ladas, A., Matthews, S. and Curry, S. Structure of tandem RNA recognition motifs from polypyrimidine tract binding protein reveals novel features of the RRM fold. (2000) *EMBO J.*, **19**, 3132-3133.

Lamichhane, R., Daubner, G. M., Thomas-Crusells, J., Auweter, S. D., Manatschal, C., Austin, K. S., Valnuik, O., Allain, F. H.-T. and Rueda, D. RNA looping by PTB: Evidence using FRET and NMR spectroscopy for a role in splicing regulation. (2010) *Proc. Natl. Acad. Sci.*, **107**, 4105-4110.

Oberstrass, F. C., Auweter, S. D., Erat, M., Hargous, Y., Henning, A., Wenter, P., Reymond, L., Amir-Ahmady, B., Pitsch, S., Black, D. L. and Allain, F. H.-T. Structure of PTB bound to RNA: Specific binding and implications for splicing regulation. (2005) *Science*, **309**, 2054-2057.

Vitali, F., Henning, A., Oberstrass, F. C., Hargous, Y., Auweter, S. D., Erat, M. and Allain, F. H-T. Structure of the two most C-terminal RNA recognition motifs of PTB using segmental isotope labeling, (2006) *EMBO*, **25**, 150-162.

Zuker, M. Mfold web server for nucleic acid folding and hybridization prediction. (2003)

Nucleic Acids Res., **31**, 3406-3415.

Chapter 3

NMR techniques used to measure protein motions

Traditional methods to measure fast motions (ps-ns)

Early spin relaxation methods focused on the determination of fast (ps-ns) timescale motions. Characterization of the SH2 protein domain fast timescale dynamics (Kay et al., 1998) was the first application of these methods to a biologically significant system, and typifies a standard protein relaxation analysis. Current renditions of these methods have been developed and optimized, and remain the standard starting point for complete protein dynamics analysis.

Briefly, data are collected that describe the extent of energy transfer of a nucleus with neighboring nuclei as well as its surrounding environment. Three measurements typically comprise this analysis: Longitudinal relaxation (R_1), transverse relaxation (R_2), and heteronuclear NOE enhancement. These data are then fit to a model that allows extraction of a generalized order parameter, S^2, that gives a measure of motion of each nucleus. S^2 ranges from zero, where motion is completely unrestricted, to one, where motion is completely restricted. Depending on the model selected, other information may be obtained as well, including an term that accounts for additional contributions to the transverse relaxation rate not fit to the model, or R_{ex}. However, even without further analysis, the raw R_1, R_2 and NOE data can provide insight into the dynamic nature of the molecule, and are thus worth discussing in more detail.

Molecular motions

Molecules are not rigid, yet the connection between molecular motions and function is largely unclear. NMR experiments are particularly well suited to study motions since they can detect the motion itself, allowing access to dynamics information for systems in equilibrium. Timescales of motions in molecules range from picoseconds (bond vibrations) to several seconds and longer (large scale structural rearrangements, catalysis), all of which can be probed using NMR spectroscopy (Figure 1). This wide range of accessible timescales, coupled with the atomic resolution possible using NMR methods, make NMR spectroscopy a powerful method for studying protein motions.

Basic spin dynamics background

For any spin ½ system, application of a static magnetic field generates one ground and one excited state, whose populations are described by the Boltzmann distribution. The z-axis is defined by the static magnetic field which aligns the nuclear dipoles, either parallel or antiparallel to the applied field. A radio-frequency pulse at the Larmour frequency, equal to the energy difference between the ground and excited states (on resonance), ω_L:

$$\omega_L = \gamma B$$

disrupts the thermal equilibrium distribution. The gyromagnetic ratio for a particular nucleus is denoted γ, and B is the strength of the static magnetic field. The direction and duration of the pulse dictate how the net magnetization is affected. The rates at which a perturbed system returns to an equilibrium state are termed the relaxation rates and are the basis for most NMR dynamics experiments.

Consider a system defined with cartesian coordinates, in the presence of a static magnetic field, and the absence of an on-resonance pulse. Each spin precesses at its Larmour frequecy and the net magnetization is aligned along the z-axis. When an on-resonance pulse is applied, the net magnetization is shifted either into the x-y plane (if the pulse is 90º or $\pi/2$) or inverted along the z-axis (for a 180º or π pulse). After the pulse is turned off, two things happen: the magnetization dephases in the x-y plane and returns to the z-axis, and the equilibrium state is eventually achieved. For a heteronuclear AX system, simple experiments can be used to measure the rates of each of these processes separately.

For ^{15}N-^{1}H and ^{13}C-^{1}H systems, the rate of return of the net magnetization to the z-axis is termed longitudinal, or spin-lattice relaxation, and is dominated by dipole-dipole and chemical shift anisotropy (CSA) relaxation mechanisms:

$$R_1 = R_1^{DD} + R_1^{CSA}$$

A simple 1D inversion recovery pulse sequence illustrates how this rate may be measured (Figure 2). Typically, two-dimensional adaptations of this, using INEPT or similar tools to transfer the magnetization from the proton to the nucleus of interest, in this case ^{15}N, and a reverse INEPT to transfer the magnetization back in order to detect in the proton dimension after the delay, are used to obtain per residue longitudinal relaxation rate information for proteins.

The rate of dephasing of spins in the x-y plane is termed transverse or spin-spin relaxation and is also dominated by dipole-dipole interactions, but also includes a

contribution from exchange processes.

$$R_2 = R_2^{DD} + R_2^{CSA} + R_{ex}$$

A simple Hahn echo one dimensional experiment illustrates how the transverse relaxation rate can be measured (Figure 3). Similar two-dimensional modifications are employed to obtain residue specific information for proteins. In practice, a similar sequence developed by Carr, Purcell, Meiboom and Gill (CPMG) is generally used. This version not only removes imperfections in the 180º pulse, but also introduces the so called CPMG pulse train, that is, multiple pulses at variable frequencies (centered at the same frequency) so that the delay between pulses can be varied while maintaining constant magnetization. This allows direct measurement of the loss of coherence, and ensures that other processes to not contribute to the decay.

Dipole-dipole interactions are an important relaxation pathway, so it should come as no surprise that dipolar coupled spins can 'sense' each other. In a dipolar coupled system, saturation of one spin will affect the equilibrium populations of the other spin. This effect is called NOE enhancement, and is measured using simple paired experiments, one with protons saturated during the evolution time, and the other without. In a covalently coupled spin system, the ratio of peak intensities gives a measure of the extent of communication between the coupled spins, where values close to unity (very little difference in the saturated and unsaturated peak intensities) imply rigidity of the bond vector, and more flexible residues give rise to NOE ratios of lower value. For a homonuclear system, NOE values range from 1 to -1. However, for a heteronuclear system, the difference in gyromagnetic ratios must be accounted for, so the while the

upper limit remains 1, very flexible residues may give rise to much lower NOE values.

Relaxation data analysis: The ModelFree approach

We can recast the information from the above measurements in a more intuitive format by fitting the data to a model and calculating parameters that describe the motions. This is most commonly done using Art Palmer's ModelFree Program, which is based on the theory developed by Lipari and Szabo (Lipari & Szabo, 1982, a,b). Assuming that relaxation is dominated by dipole-dipole and CSA effects, the relaxation parameters depend on the value of the spectral density function, J(ω), at five characteristic frequencies. Because it is difficult, if not impossible to explicitly determine the spectral densities at all five frequencies, a simpler model is employed that describes the relaxation using a small number of physically meaningful parameters (Palmer, et al 1991). For an isotropically tumbling molecule, the model is:

$$J(\omega) = \frac{2}{5}\left[\frac{S^2\tau_M}{1+(\omega\tau_M)^2} + \frac{(1-S^2)\tau}{1+(\omega\tau)^2}\right]$$

and can be extended to account for axially symmetric tumbling. This is the Model-Free approach originally proposed by Lipari and Szabo in 1982, and contains a generalized order parameter S^2, which provides a measure of spatial restriction of the bond vector from zero, for isotropic motion, to unity for no motion in a fixed frame of reference. τ is an effective correlation time, where

$$\tau^{-1} = \tau_M^{-1} + \tau_e^{-1}$$

τ_M is the overall tumbling time for the molecule, or the time it takes the molecule to rotate by one radian, and τ_e is the local correlation time of the individual bond vector.

The Model-Free nomenclature arises from the fact that τ_M and S^2 are model-independent parameters. Once these values have been determined, the results are interpreted within the framework of a model. Because fitting to extract the parameters that describe internal motion is influenced by the overall tumbling of the molecule, it is important to choose a model that accurately describes the way the molecule tumbles in solution. Inaccurate modeling of molecular tumbling commonly results in low magnitude R_{ex} terms throughout the protein as the program is unable to fit accurately to fast timescale dynamics.

Descriptions of the R_1, R_2 and NOE relaxation for a heteronuclear system were derived by Abragam in 1961, considering only the dipole-dipole and CSA contributions to relaxation. These expressions form the basis of the ModelFree analysis, and are given below:

$$R_1 = (d^2/4)[J(\omega_H-\omega_X) + 3J(\omega_X) + 6J(\omega_H+\omega_X)] + c^2J(\omega_X)$$

$$R_2 = (d^2/8)[4J(0) + J(\omega_H-\omega_X) + 3J(\omega_X) + 6J(\omega_H) + 6J(\omega_H+\omega_X)] + (c^2/6)[4J(0) + 3J(\omega_X)] + R_{ex}$$

$$NOE = 1 + (d^2/4R_1)(\gamma_X/\gamma_H)[6J(\omega_H+\omega_X) - J(\omega_H-\omega_X)]$$

where $d = \mu_0 h\gamma_X\gamma_H(r_{XH}^{-3})/(8\pi^2)$, $c = \omega_X\Delta\sigma/(3^{-1/2})$, μ_0 is the permeativity of free space, h is Plank's constant, ω_H and ω_X are the Larmour frequencies of the ^1H and X (either ^{13}C or ^{15}N) spins, respectively, r_{XH} is the X-H bond length, and $\Delta\sigma$ is the chemical shift

anisotropy of the X spin, assumed to be axially symmetric with an axis collinear with the X-H bond vector. The R_{ex} term in the expression for R_2 is purely phenomenological, included to account for exchange processes that contribute to the transverse relaxation rate. Using the above expressions along with an appropriate form of the spectral density function (as determined by the model selection procedure discussed below), parameters that describe residue specific motion on a fast (ps-ns) time scale can be extracted. In addition, an R_{ex} term may be added to account for slow (μs-ms) motions if needed.

ModelFree analysis typically involves reiteration of three steps: initial estimation of the tumbling time of the molecule, model selection, and final optimization. Depending on the structural information available, there are several adequate methods to estimate the tumbling time of the molecule. For this work, experimental R_1 and R_2 data were used according to the method of Tjandra et al. (1995) using the R2R1_diffusion software available on the Palmer group website. This software estimates diffusion tensors for spherical or axially symmetric molecules and is designed to be compatible with the ModelFree spectral density functions.

Alternately, a trimmed mean of the R_2/R_1 ratio can be used to estimate the tumbling time of a molecule, excluding the residues which are suspected to have exchange contributions to relaxation. However, for very flexible proteins, such as the proteins studied here, this method is not appropriate since most amides will have a transverse relaxation rate that includes contributions from exchange processes. Also, because R_{ex} terms can arise as a result of a poorly defined diffusion tensor, it is prudent to

obtain the most accurate description possible.

Once a diffusion tensor is estimated, one may proceed to the model selection step. Each model uses different parameters to describe the molecular motions, increasing in complexity as necessary to fit the data. The available models are:

Model 1: S^2

Model 2: S^2 and τ_e

Model 3: S^2 and R_{ex}

Model 4: S^2, τ_e and R_{ex}

Model 5: S_f^2, S_s^2 and τ_e

In model 5, the ModelFree formalism has been extended to contain both fast (S_f^2) and slow (S_s^2) generalized order parameters (Clore et al, 1990). Importantly, data must be collected at multiple static magnetic field strengths if more than two modelfree parameters (the tumbling time, τ_M is also included in the fit) are to be used in the fitting. The algorithm for model selection proceeds as described in Mandel et al (1995), and is summarized in figure 4. Briefly, attempts are made to fit the data to each model, in order of increasing complexity, until an adequate fit has been obtained. Implementation of more complex models is justified in ModelFree via a F-statistic test (see Mandel et al (1995)).

The final step in the ModelFree analysis is the simultaneous optimization of the overall rotational diffusion model and the internal motional parameters for each spin. Generally, several iterations of the three steps are necessary until a self-convergent model of the system is obtained.

Interpretation of ModelFree results

The results of the ModelFree analysis provide residue specific motional information about a given system. An accurate interpretation of the results, however, requires revisitation of some of the limits of the method. First of all, the analysis is based on theoretical expressions that take into account only dipole-dipole and CSA relaxation mechanisms. Autorelaxation (R_0) effects are not taken into account at all, and R_{ex} terms provide some indication of slow time scale chemical exchange, but are not measured directly.

Experience shows that an inability to converge is not the only way this method can fail. Theoretically, the generalized order parameter, S^2, has an upper limit of 0.95, nonetheless, the program will sometimes return values up to 1.0. This is clearly a nonsensical result, and is usually, but not always, avoided by properly selecting the initial diffusion tensor estimation. Suggestions have been made to expand to possible range of values to 1.2 so that the fits don't get stuck at 1.0, but this strategy failed to fix the problem for the proteins studied here. Spins which consistently yield order parameters which are too high are unable to be adequately fit, and are thus discarded.

It is well known that a poorly described diffusion tensor can lead to artefactual R_{ex} terms. In this case, it is important to make sure that the tumbling time obtained does not indicate dimer, or other oligomer formation, or non-specific associations, that may erroneously lead to R_{ex} terms. In addition, it is imperative to use at least two data sets, collected at different static magnetic field strengths, to ensure that the R_{ex} terms do not arise from underfitting (ie: there are more input paramters than output parameters). However, even if the R_{ex} terms do not arise from some error, ModelFree scales the magnitude of the terms quadratically with static magnetic field strength. This approximation, however, only holds for rigid molecules with a two-state exchange, where one state is much more highly populated than the other state, and the exchange is fast on the chemical shift timescale (Millet et al. 2000). None of these criteria are expected to hold *a priori* for the systems studied here, thus, for the work described herein, ModelFree can identify residues which undergo slow exchange, but will not yield accurate magnitudes of the exchange terms.

Since the tumbling time of the molecule is an integral component of the ModelFree analysis, it is a good idea to verify that this result is of an acceptable value. To this end, several methods were employed to independently measure the tumbling time of the protein constructs used in this study. The Stokes-Einstein relationship was used to calculate the tumbling times, assuming spherical molecules. The tumbling time was also calculated considering the structures of the constructs using HYDRONMR. Finally, TRACT NMR experiments were performed to experimentally measure the tumbling times. These techniques will be discussed in detail in the next chapter.

Direct Measurement of slow exchange

The most satisfying way to confirm the presence of slow motions, however, is to measure the motions directly. NMR methods have been developed that allow direct detection of microsecond to millisecond motion. It helps to have an idea of the timescale of the motions of interest in order to choose an appropriate experiment, but is not necessary. CPMG based dispersion experiments access millisecond timescales, and are the least time-consuming of the methods discussed here. $R_{1\rho}$ dispersion experiments access microsecond timescales, but have significantly more technical difficulty and require much longer acquisition times. Finally, relaxation interference experiments offer an alternative approach that enables confirmation of residues that have slow exchange contributions to relaxation, but lack the quantitative power of the former techniques.

Measurement of millisecond motions: CPMG dispersion

The exchange contribution to transverse relaxation was first formally accounted for by modification of the phenomenological equations for bulk magnetization known as the Bloch equations by Hahn and Maxwell (1952) and McConnell (1958). These ideas were expanded and applied to a practical system in the classic paper by Carver and Richards (1972), where a general solution for two-site chemical exchange was described from the dependence of the transverse relaxation rate on the CPMG pulse separation, τ_{cp}. Briefly, for a two-state system:

$$A \leftrightarrow B$$

with forward reaction rate k_A and reverse reaction rate k_B, the rate of exchange between states A and B is given as $k_{ex} = k_A + k_B$. The populations of each state are given as p_A

and p_B, and the frequency difference between the two states is denoted $\Delta\omega = \omega_A - \omega_B$.

During the experiment, the CPMG pulse train (a series of 180º pulses separated by time τ_{cp}) focuses spins that are not involved in exchange processes. Spins which exchange on the timescale of the pulsing, spend time in more than a single state, and thus are not refocused, so the peak intensity recovered from exchanging spins is less than that for non-exchanging spins. As the time between 180º pulses, τ_{cp} increases, the chance that a spin will sample more than a single environment increases, and the effective decrease in peak intensity is enhanced. For spins undergoing exchange on the timescale of the CPMG pulse separation, a plot of the effective transverse relaxation rate, R_2^{eff}, dependence on τ_{cp}, termed the dispersion profile, yields a curve that can be fit by the following expression (Downing 2004):

$$R^2(\tau_{cp}^{-1}) = \frac{1}{2}\left(R_{2A}^0 + R_{2B}^0 + k_{ex} - \tau_{cp}^{-1}\cosh^{-1}\left[D_+\cosh(\eta_+) - D_-\cos(\eta_-)\right]\right)$$

where:

$$D_\pm = \frac{1}{2}\left[\pm 1 + \frac{\psi + 2\Delta\omega^2}{(\psi^2 + \xi^2)^{1/2}}\right]$$

$$\eta_\pm = \frac{\tau_{cp}}{\sqrt{2}}\left[\pm\psi + (\psi^2 + \xi^2)^{1/2}\right]^{1/2}$$

$$\psi = (R_{2A}^0 - R_{2B}^0 - p_A k_{ex} + p_B k_{ex})^2 - \Delta\omega^2 + 4 p_A p_B k_{ex}^2$$

$$\xi = 2\Delta\omega(R_{2A}^0 - R_{2B}^0 - p_A k_{ex} + p_B k_{ex})$$

Depending on the system, certain assumptions can be made in order to simplify the above expression. As long as $p_A >> p_B$, exchange on any timescale can be described as (Downing 2004):

$$R_2(\tau_{cp}^{-1}) = R_2(\tau_{cp}^{-1} \to \infty) + \frac{p_A p_B \Delta\omega^2 k_{ex}}{k_{ex}^2 + (p_A^2 \Delta\omega^4 + 144\tau_{cp}^{-4})^{1/2}}$$

Where $R_2(\tau_{cp}^{-1}) \to \infty$ can be approximated by the measured R_2. If it is not practical to assume $p_A >> p_B$, then the following expression may be used as long as the system is in fast exchange (Downing 2004).

$$R_2(\tau_{cp}^{-1}) = R_2(\tau_{cp}^{-1} \to \infty) + (p_A p_B \Delta\omega^2 / k_{ex}) \left[1 - \frac{2\tanh(k_{ex}\tau_{cp}/2)}{k_{ex}\tau_{cp}} \right]$$

Regardless of the expression used, the primary goal is to extract kinetic information, k_{ex}, as well as thermodynamic information, p_A, p_B, and structural information, $\Delta\omega$. One of the main limitations of this method, however, is that the theory has only been worked out for a simple two-state system. If the system is suspected to have more than two states, as is likely for the system described herein, it is not yet possible to extract such information. Even if fitting is not appropriate, the strategy still has some value, in that it can be used to identify residues that undergo exchange on a millisecond timescale.

Using this method, no amides in PTB1:34 were found to exchange on a millisecond timescale. However, some amides had elevated R_2^{eff} values that did not decay, suggesting that these residues are in motion just slightly faster than can be detected using this method. Decreasing the temperature to 10ºC did not lead to detectable dispersion for these residues, but did provide a clue that at least some amides in the protein are in exchange in a slightly faster regime.

Measurement of microsecond motions: $R_{1\rho}$ dispersion

$R_{1\rho}$ experiments allow access to microsecond motions using methods analogous to CPMG dispersion experiments. In both cases, the contribution of exchange to the transverse relaxation rate is varied as a function of applied field strength while the magnetization is in the x-y plane (for CPMG) or in the tilted plane (for $R_{1\rho}$). For CPMG experiments, the applied field strength is defined as the spacing between 180º pulses in the CPMG pulse train. For $R_{1\rho}$ experiments, the applied field is continuous, increasing in strength through either increased power or position in the rotating frame. Higher field strengths are able to detect faster exchange. Thus, practical limitations, ie: not frying the sample or probe, determine the fastest motions that can be determined using this method.

The ability to measure fast exchange rates in the rotating frame was first noted by Deverell, et al. (1970) using cyclohexane chair to chair isomerization as an example. This method could be used to extract exchange rates, but was limited to the case where $p_A = p_B$. The theory was expanded by Davis et al. (1994), based on the Carver-Richards equations to obtain a general solution for a two state system that allowed extraction of the

same kinetic and thermodynamic parameters as a CPMG dispersion analysis. Assuming fast exchange, the transverse relaxation rate in the rotating frame can be expressed as:

$$R_{1\rho} = R_1 \cos^2\theta + R_2 \sin^2\theta + p_A p_B (\Delta\omega^2) \frac{k_{ex}}{k_{ex}^2 + \omega_e^2} \sin^2\theta$$

Here, θ describes the angle between the effective magnetic field and the z-axis:

$$\theta = \tan^{-1} \frac{\omega_1}{\delta\omega}$$

where δω is the frequency difference between the applied field and the resonance frequency of the spin of interest. For an off-resonance experiment, the effective spin-locking field, ω_e, is described by the vector sum of of the spin lock field, ω_1 and δω:

$$\omega_e = \sqrt{\omega_1^2 + \delta\omega^2}$$

In the case where δω = 0, that is, the applied field is on-resonance, the above expression simplifies to:

$$R_{1\rho} = R_2 + p_A p_B (\Delta\omega^2) \frac{k_{ex}}{k_{ex}^2 + \omega_1^2}$$

$R_{1\rho}$ relaxation rates can be measured either off-resonance, where the applied field is varied by moving the frequency away from the resonance frequency of the spin of interest, or on-resonance, where the applied field is varied by increasing the RF power of the pulse. Off-resonance methods can access a larger range of motions since the field is increased by positioning rather than increased power, so that sample and probe heating is

not as much of a problem. However, the implementation and analysis of off-resonance methods is much more complicated than that of on-resonance methods, requiring not only a separate experiment to calibrate the effective spin-lock field strength, ω_e, but also a more complicated fitting function. In addition, the latest versions of these pulse sequences were written for Brüker spectrometers, and have not been adapted for the Varian spectrometers that are available here.

On-resonance $R_{1\rho}$ experiments offer much simpler analysis and a straightforward determination of applied field strength. In addition, the standard Varian biopack $R_{1\rho}$ experiment could be modified to allow user control of the RF field strength. However, near-resonance may be a better description of these experiments since it is not practical to run a separate dispersion profile for each amide resonance. Instead, strips of 5 ppm in the nitrogen dimension were used to approximate the resonance frequency of all amides in that region. The transverse relaxation rate in the rotating frame for amides in exchange on a microsecond timescale is dependent on the spin lock field strength. Measurement of this rate at varying field strength is expected to generate a dispersion profile that can be fit as a decay curve with the simplified $R_{1\rho}$ dispersion expression, and the same kinetic and thermodynamic parameters as in the CPMG dispersion analysis can be extracted.

This method, however, also suffers from the same limitations as the CPMG dispersion analysis in that the dispersion curve can only be fit by a simple two-state model. For this system, which is not two-state, quantitative descriptions of the motion are not possible. Since every point on the dispersion curve requires five data sets

(corresponding to 5 strips of 5 ppm each), the analysis was simplified by definition of $\Delta R_{1\rho}$, the difference in $R_{1\rho}$ values at two different spin lock field strengths. Since our modified pulse sequences had never been used before, the powers of the two spin lock field strengths were chosen conservatively in order to maximize the difference of $R_{1\rho}$ rates, yet to stay well within the safe power range of the instrument.

A simple scheme of this method is depicted in figure 5, and illustrates how, since a fit is not used to extrapolate to the slow pulsing limit of the experiment, the overall magnitude of the exchange contribution to the transverse relaxation rate will be underestimated. Nonetheless, this method directly monitors microsecond motions, and will thus robustly identify any residues that have exchange on this timescale. While some residues may be missed, the method has the advantage of minimizing the possibility of false positives, making it a good complement to the other methods used. The major advantage of this method is that is does not depend on fitting or assumptions about the system. Microsecond dynamics are the only processes that will give rise to a positive result with this experiment.

Identification of exchanging residues: Relaxation interference experiments

The phenomenon of interference between N-H dipolar coupling and CSA relaxation mechanisms and its contribution to transverse relaxation has been acknowledged since the 1950's. More than three decades later a formalism for describing a scalar coupled heteronuclear system, with both nuclei spin ½, that relaxes via internuclear dipolar and anisotropic chemical shift interactions was described by

Goldman (1984). The value of this relaxation interference effect is that it allows for the measurement of transverse relaxation independent of exchange, or the so-called exchange-free transverse relaxation rate, η_{xy}.

$$\eta_{xy} = R_2^{DD} + R_2^{CSA}$$

Comparison of this rate with the traditional transverse relaxation rate, R_2 (which includes contributions from exchange), facilitates a robust identification of residues involved in exchange.

The observation that the two peaks of ^{15}N-amide proton doublets have different line widths arises from the different relaxation rates of each state. The relaxation rates for each the α and β spin states, respectively, are given as:

$$R_\alpha = \lambda + \eta$$

$$R_\beta = \lambda - \eta$$

Where λ represents the auto-correlated relaxation processes and η represents the cross-correlated relaxation processes that contribute to transverse relaxation. Thus, the difference of the two rates,

$$R_\alpha - R_\beta = \eta_{xy}$$

yields the exchange-free transverse relaxation rate. Residues which have significant exchange contribution to the transverse relaxation rate are identified as those residues that have R_{ex} greater than zero, where:

$$R_{ex} = R_2 - \kappa\eta_{xy}$$

In this case, R_2 is the traditional transverse relaxation rate measured initially, and κ is a constant that accounts for autorelaxation effects. Determination of κ is not trivial and

will be discussed in the next section.

In 1996, a pair of pulse sequences to measure the transverse relaxation rates of each of the doublet components of ^{15}N-labeled protein backbone amides was reported (Tjandra et al, 1996). In 1998 the method was expanded to include a similar method for analyzing longitudinal relaxation rates. A major limitation of these methods was that the scaling between the two measurements was difficult, yet critical for an accurate analysis. Liu and Prestegard (Liu & Prestegard, 2008) addressed this problem by further improving the method so that both rates could be measured in a single experiment, eliminating the scaling problem.

Other potential sources of error involve the assumption that the principle axes of the dipole-dipole interaction and the ^{15}N CSA are co-linear. Also, the nitrogen CSA is assumed to be axially symmetric. However, deviations from these assumptions have been found to impact the measured rates by less than 10% (Tjandra, et al. 1996). In addition, cross-correlation between ^{15}N CSA and the dipole-dipole interactions with distant protons may lead to some degree of error. However, the pulse sequence design of Liu and Prestegard limits this effect to an approximate underestimation of 0.75%. If greater accuracy is desired, perdeuterating the protein would further diminish this effect, however, the degree of uncertainty from the assumptions discussed above render this amount of error negligible.

Determination of κ

As mentioned above, accurate determination of R_{ex} using relaxation interference

experiments requires that autorelaxation effects be taken into account. In general, κ is a constant over all residues, under the assumption that the ^{1}H-^{15}N dipolar and the ^{15}N CSA axes are co-linear. The traditionally recommended method for determination of κ is to take a trimmed mean of R_2/η_{xy} for all residues not involved in exchange processes (Wang et al., 2003). This method seems appropriate for proteins that are thought to be mostly rigid, with only a few residues undergoing exchange and thus was used to determine κ for RRM4. The RRM4 measured R_2/η_{xy} (at 700MHz) for residues not undergoing exchange was found to be 1.35.

κ is independent of any assumptions involving local or overall motions, and so can also be calculated as a theoretical ratio of dipolar and CSA relaxation effects. This method is preferred for systems that are thought to have most residues involved in exchange. Since PTB1:34 is larger, and does not have a clear baseline of residues which do not undergo exchange, its κ was calculated as the theoretical ratio of R_2/η_{xy} for dipolar and CSA relaxation pathways (Fushman et al., 1998):

$$\kappa = \frac{\delta^2 + p^2}{2\delta p}$$

Where δ is the chemical shift anisotropy of the ^{15}N nucleus,

$$\delta = \frac{\gamma_N B_0 \Delta\delta_N}{3\sqrt{2}}$$

And p is the $^{1}H/^{15}N$ dipole-dipole coupling:

$$p = \frac{\mu_0 \gamma_H \gamma_N h}{16\pi^2 \sqrt{2} r_{HN}^3}$$

Where γ_N and γ_H are the ^{1}H and ^{15}N gyromagnetic ratios, B_0 is the static magnetic field strength, $\Delta\delta_N$ is the span of the principal components of the ^{15}N chemical shift tensor, μ_0 is the permeativity of free space, h is Planck's constant, and r_{HN} is the $^{1}H/^{15}N$ internuclear distance (1.02 Angstroms). At 700MHz, this value was determined to be 1.06.

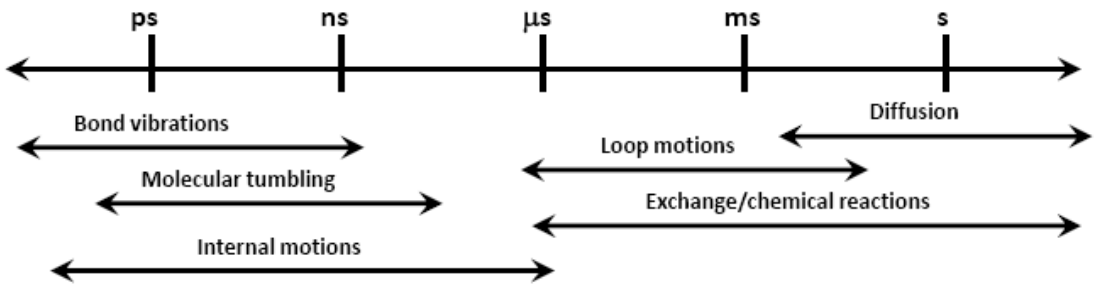


Figure 1: Motions are important for biological molecules on many different timescales. Fast motions such as bond vibrations, molecular tumbling and internal motions usually occur in the range of picoseconds to nanoseconds. Loop motions, molecular diffusion, as well as chemical and conformational exchange generally are much slower, in the realm of microsecond to millisecond, and even seconds. NMR methods are particularly well suited for studying protein motions since they can access motions over this entire range of timescales.

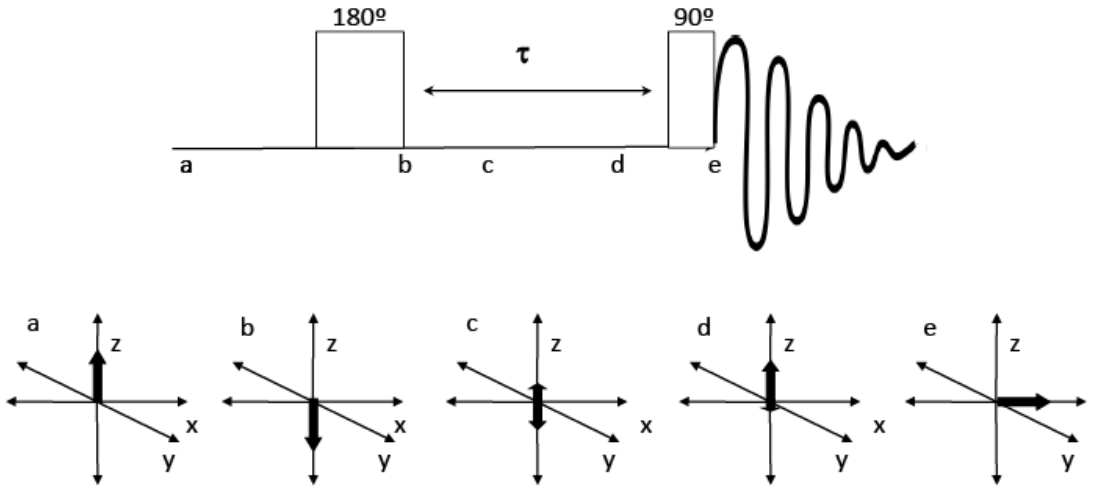


Figure 2: Simple schematic diagram of a one dimensional inversion recovery NMR experiment used to measure longitudinal relaxation rates. The top panel shows the pulse sequence, and the bottom panel depicts the net magnetization in cartesian coordinates at specified points of the pulse sequence. At point a, prior to pulsing, the net magnetization is aligned with the B_0 field along the z-axis. At point b, a 180º radio frequency pulse is applied at the Larmour frequency that flips the net magnetization to the -z-axis. During the delay time, τ, the system relaxes, and the net magnetization shifts back to its equilibrium state, aligned with the z-axis. Points c and d indicate the return to the equilibrium state as the relaxation time, τ, progresses. Immediately following τ, a 90º pulse (-y) is applied that flips any magnetization that has returned to the z-axis to the x-axis for detection (point e), and the FID is collected. Typically, peak intensities are recorded at several values of τ, and an exponential decay of signal intensity as a function of τ describes the longitudinal relaxation rate.

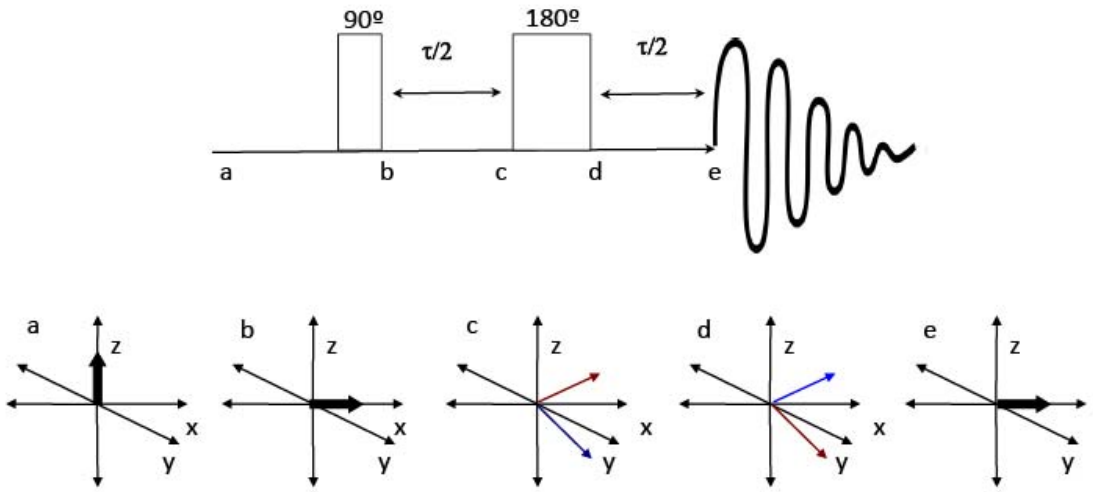


Figure 3: Simple schematic diagram of a one dimensional Hahn-echo NMR experiment used to measure transverse relaxation rates. The top panel shows the pulse sequence, and the bottom panel depicts the net magnetization in cartesian coordinates at specified points in the pulse sequence. At point a, prior to pulsing, the net magnetization is aligned with B_0 along the z-axis. At point b a 90º pulse has been applied that flips the magnetization to the x-axis. During the first τ/2 delay, the magnetization dephases in the x-y plane. For simplicity, only two spins are shown with red and blue arrows at point c, however in reality, all spins fan out in the x-y plane during this delay. At point d, a 180º pulse along the y-axis reverses the position of each spin in the x-y plane. After an identical delay time, τ/2, the spins, now moving in the opposite direction as a result of the 180º pulse, reconvene at the x-axis, and the FID is collected. As τ increases, the amount of magnetization that is refocused is reduced, and peak intensities diminish. Peak intensities are recorded for several values of τ, and the decay is fit to an exponential function to extract the transverse relaxation rate.

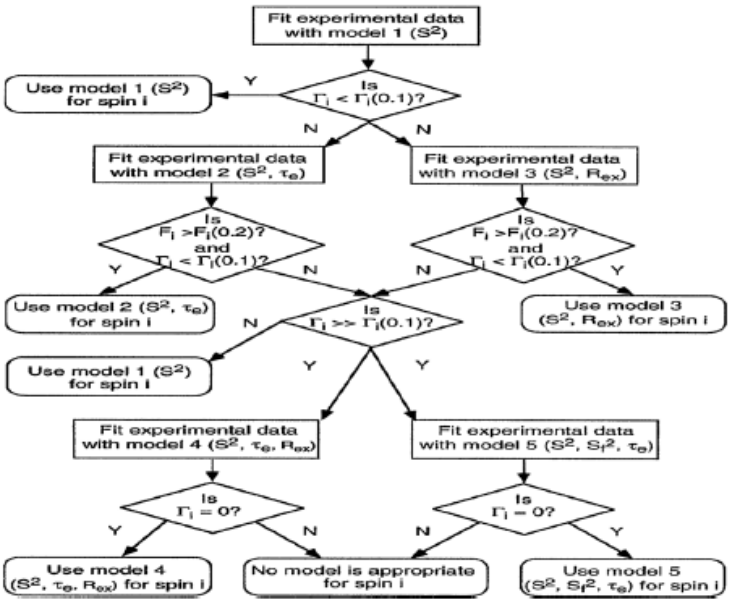


Figure 4: Flowchart of the model selection strategy used by ModelFree. The use of additional parameters for fitting is justified by comparison of simulated distributions of Γ and F for each spin, i. Adapted from Mandel et al. (1995).

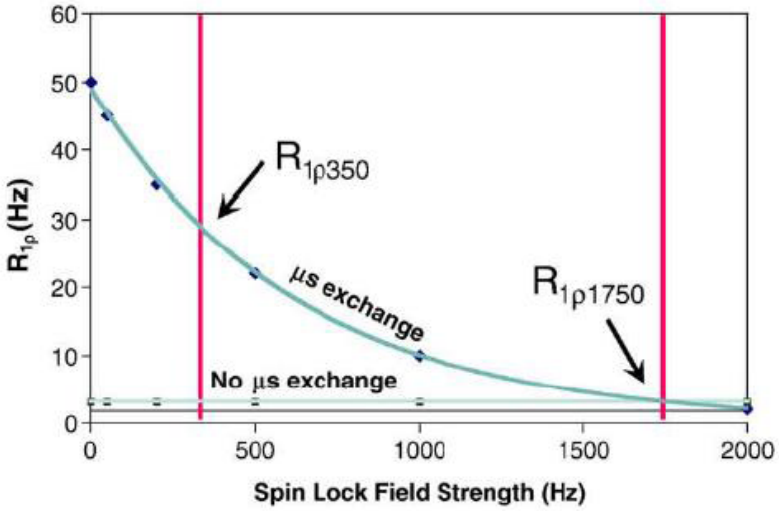


Figure 5: A schematic representation of the determination of $\Delta R_{1\rho}$. Examples are shown for dispersion curves for a residue with microsecond exchange (blue curve) and a residue with no microsecond exchange (grey curve). Importantly, this figure shows the propensity of the method to underestimate the exchange contribution to R_2 as the $R_{1\rho}$ rates for the lowest spin lock strength values are not accessible since a fit is not used to extrapolate to a spin lock field strength of 0 Hz.

References:

Abragam, A. Principles of nuclear magnetism. (1961). Oxford, Claredon Press.

Carver, J. P., & Richards, R. E. A general two-site solution for the chemical exchange produced dependence of T_2 upon the Carr-Purcell pulse separation. (1972) *J. Magn. Reson.*, **6**, 89-105.

Clore, G. M., Szabo, A., Bax, A., Kay, L. E., Driscoll, P. C., and Gronenborn, A. M. Deviations from the simple two-parameter Model-Free approach to the interpretation of nitrogen-15 nuclear magnetic relaxation of proteins. (1990) *J. Am. Chem. Soc.*, **112**, 4989-4991.

Davis, E. G., Perlman, M. E., & London, R. E. Direct Measurements of the dissociation-rate constant for inhibitor-enzyme complexes via the T1r and T2 (CPMG) methods. (1994), *J. Magn. Reson. B.*, **104**, 266-275.

Deverell, C., Morgan, R. E., & Strange, J. H. Studies of chemical exchange by nuclear magnetic relaxation in the rotating frame. (1970) *Mol. Phys.*, **18**, 553-559.

Downing, A. K. Protein NMR techniques (2004) Humana Press, Inc.

Fushman, D., Tjandra, N., & Cowburn, D. Direct measurement of ^{15}N chemical shift anisotropy in solution. (1998) *J. Am. Chem. Soc.*, **120**, 10947-10952.

Goldman, M. Interference effects in the relaxation of a pair of unlike spin-1/2 nuclei. (1984) *J. Magn. Reson.*, **60**, 437-452.

Hahn, E. L., & Maxwell, D. E. Spin echo measurements of nuclear spin coupling in molecules. (1952) *Phys. Rev.*, **88**, 1070-1086.

Kay, L. E., Torchia, D. A., Bax, A. Backbone dynamics of proteins as studied by 15N inverse detected heteronuclear NMR spectroscopy: Application to staphylococcal

nuclease. (1989) *Biochemistry*, **28**, 8972-8979.

Lipari, G., Szabo, A. Model-Free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity. (1982) *J. Am. Chem. Soc.*, **104**, 4546-4559.

Lipari, G., Szabo, A. Model-Free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 2. Analysis of experimental results. (1982) *J. Am. Chem. Soc.*, **104**, 4559-4570.

Liu, Y. & Prestegard, J. H. Direct measurement of dipole-dipole/CSA cross-correlated relaxation by a constant-time experiment. (2008) *J. Magn. Reson.*, **193**, 23-31.

Mandel, A. M., Akke, M., and Palmer, A. G. 3rd. Backbone dynamics of *Escherichia coli* ribonuclease HI: Correlations with structure and function in an active enzyme. (1995) *J. Mol. Biol.*, **246**, 144-163.

McConnell, H. M. Reaction rates by nuclear magnetic resonance. (1958) *J. Chem. Phys.*, **28**, 430-431.

Millet, O., Loria, J. P., Kroenke, C. D., Pons, M., Palmer, A. G. 3rd. The static magnetic field dependence of chemical exchange lineboradening defines the NMR chemical shift time scale. (2000) *J. Am. Chem. Soc.*, **122**, 2867-2877.

Palmer, A. G. 3rd, Rance, M., Wright, P. E. Intramolecular motions of a zinc finger DNA-binding domain from Xfin characterized by proton-detected natural abundance 13C heteronuclear NMR spectroscopy. (1991) *J. Am. Chem. Soc.*, **113**, 4371-4380

Tjandra, N., Feller, S. E., Pastor, R. W., Bax, Ad. Rotational diffusion anisotropy of human ubiquitin from ^{15}N NMR relaxation data. (1995) *J. Am. Chem. Soc.*, **117**, 12562-12566.

Tjandra, N., Szabo, A. & Bax, A. Protein backbone dynamics and ^{15}N chemical shift anisotropy from quantitative measurement of relaxation interference effects. (1996) *J. Am. Chem. Soc.*, **118**, 6986-6991.

Wang, C. Y., Rance, M., & Palmer, A. G. Mapping chemical exchange in proteins with MW > 50kDa. (2003) *J. Am. Chem. Soc.*, **125**, 8968-8969.

Chapter 4

Methods for molecular tumbling time measurement

Knowledge of the tumbling behavior of molecules in solution is requisite to the interpretation of the types of analysis used in the studies described herein. Protein oligomerization as well as non-specific or transient protein associations may influence the outcome of NMR relaxation experiments, and may also interfere with RNA binding studies. The ability to accurately analyze protein motions using ModelFree also depends on an accurate description of how the molecule tumbles in solution. For the types of methods used here, this is typically done using the metric of tumbling time, τ_c, the time it takes the molecule to tumble one radian in solution. This value is included in the spectral density function, so any NMR method that relies on spectral densities to describe motions is only as accurate as the estimation of tumbling time. As discussed in the previous chapter, the ModelFree program uses an initial estimate of the tumbling time, usually based on a trimmed mean of the relaxation data, and then uses this estimate as a starting point for fitting, where the tumbling time is further refined. The rate at which the molecule tumbles in solution is reflected in NMR relaxation rates, so protein-protein interactions in solution, even if transient and non-specific, may lead to artificially long tumbling times.

The shape of the molecule is also important for determining its tumbling time. Spectral densities for molecules that are rigid, and spherical or axially symmetric are established and included in the ModelFree analysis. However, molecules that tumble anisotropically, or are very flexible, may not be described accurately using this method.

Thus, it is important to verify that the tumbling times extracted by ModelFree make sense. Although ModelFree determines the tumbling time based on structural coordinates (for an axially symmetric system), experimental data, as well as other fitting parameters, depending on the model that is used, other methods do not use all of the same parameters. Furthermore, there are no methods that can reliably account for anisotropic tumbling and molecular flexibility, both of which likely play a role in the system studied here. However, using multiple methods, keeping in mind the limitations of each method, we should be able to get a handle on an acceptable range of values where the tumbling time is likely to be accurate. The methods used here include a simple calculation using the Debye-Stokes relationship, a calculation which is based on the structure of the molecule called HYDRONMR, and an experimental NMR method TRACT (Trosy for rotational correlation times).

Estimation of tumbling time based on molecular mass

For a spherical molecule moving through an aqueous solution, the tumbling time, τ_c, is given by the Debye-Stokes relationship:

$$\tau_c = \frac{4\pi r^3 \eta}{3kT}$$

where k is Boltzmann's constant, T is the temperature, η is the viscosity, and r is the radius of a sphere estimated from the molecular mass of each protein construct. Clearly this method does not account for molecular shapes that deviate from a perfect sphere or molecular flexibility that may impact tumbling time. The advantage of this method,

however, is that its simplicity ensures that the results are not impacted by experimental artifacts or theoretical assumptions that could introduce large errors. So, while the results of this calculation may not be entirely accurate, they will likely not be too far from the truth. The results obtained for RRM3, RRM4 and PTB1:34 are given in Table 1.

Estimation of tumbling time based on HYDRONMR calculations

Although most proteins are not perfectly spherical, estimating the tumbling time for non-spherical proteins has been a long-standing challenge. Methods such as ModelFree are able to account for axially symmetric molecules, however, no generally useful theory had been developed for determining the tumbling times of fully anisotropic molecules. This problem was addressed by the introduction of HYDRONMR by Garcia de al Torre, et al. (2000). This program uses atomic coordinates obtained from a PDB file. Hydrodynamic beads are packed onto the protein backbone, and from this, the program builds an appropriate hydrodynamic model of the protein, computes a fully anisotropic rotational diffusion tensor and determines a tumbling time. In addition, HYDRONMR uses the PDB coordinates to extract the bond vectors involved in dipolar relaxation, and uses this information to estimate the NMR relaxation parameters R_1, R_2 and $^{1}H/^{15}N$-NOE for each backbone amide.

The clear advantage of this method is that tumbling times for any shape of molecule can be determined, and deviations from spherical or axially symmetric models do not diminish the accuracy of the prediction. However, this theory was developed for small, rigid molecules, and though it has been tested successfully on over 800 proteins

from the PDB (Ryabov et al. 2006), it fails for the flexible molecules studied here. Since no experimental data are used, and the protein is assumed to be rigid, loops and tails that extend from the protein core may potentially lead to a significant overestimation of the tumbling time. The NMR structure of PTB1:34 reports that only about a third of the protein exists as secondary structure (Vitali et al. 2006). This means that the remaining two thirds exists as loops and tails that potentially lead to tumbling times that are erroneously large.

HYDRONMR analysis was completed for RRM3, RRM4 and PTB1:34 using coordinates from PDB ID:2EVZ (Vitali et al., 2006). All calculations were performed at 293 K and 0.01 poise, using 3.1 Angstrom radius for atomic elements. While the tumbling time determined for RRM4, a rigid molecule with few loops and tails extending from the protein core, were in good agreement with the times obtained from ModelFree analysis and Debye-Stokes calculation, the values obtained for RRM3 and PTB1:34 were not (Table 1). For RRM3, the calculated tumbling time of 7.5 ns is slightly larger than the value obtained from ModelFree or the Debye-Stokes calculation, an expected outcome considering the greater abundance of loops compared to RRM4. The nonsensical result of 32.3 ns for the tumbling time of PTB1:34 illustrates that this method roundly fails for this protein. Not only does PTB1:34 contain many loops and tails extending from the protein core, but we have found that this protein is highly dynamic. Both of these characteristics make the system ill-suited to be studied by this method and are expected to cause a large overestimation of tumbling time.

In addition to the extraction of tumbling times, HYDRONMR also back calculates the NMR relaxation parameters R_1, R_2 and $^1H/^{15}N$-NOE. This aspect of the program allows a useful check on the system, provided reasonable tumbling times are determined. It is important to note, however, that these calculations assume that the molecule is rigid, and also only account for dipolar relaxation mechanisms, thus any flexibility of the molecule (which contributes to R_2), or relaxation contributions from ^{15}N CSA (which is reflected in all three relaxation parameters (though to a lesser extent than dipolar effects at high field), will cause errors in these predictions as well. Comparison to experimental R_1, R_2 and $^1H/^{15}N$-NOE values measured at 500MHz is a useful indicator of the accuracy of this method for the proteins studied here.

R_1, R_2 and $^1H/^{15}N$-NOE results for the individual domains are shown in Figure 1 with experimental values shown as orange diamonds and HYDRONMR predictions shown as blue triangles. Although the HYDRONMR predicted $^1H/^{15}N$-NOE values greatly underestimate the flexibility of the molecule, the R_1 and R_2 values are in the same range as the experimentally determined values. As expected, the HYDRONMR results do not show local regions of flexibility, and variation of relaxation rates throughout the proteins is masked.

A similar comparison for PTB1:34 is an illustration of what can happen when this method completely fails (Figure 2). Not only are the most flexible regions of the protein missed, and local variability underestimated, but the R_1 and R_2 rates are drastically different from those obtained via experiment. As with RRM3 and RRM4, the $^1H/^{15}N$-

NOE predictions grossly underestimate the flexibility of the molecule.

While HYDRONMR does not seem well-suited to calculate the tumbling times of the protein constructs studied here, it does provide some clues pertaining to the nature of the molecules studied. Specifically, that this method, which is designed for rigid molecules, fails for PTB1:34, suggests that this molecule is highly dynamic, yet produces reasonable estimates for RRM3 and RRM4 tumbling times, provides a clue that these molecules are more rigid than PTB1:34.

Estimation of tumbling time using TRACT experiments

Because of the importance of molecular tumbling time not only in NMR data analysis, but also to ensure that the molecules are not forming oligomers or non-specific aggregates, TRACT experiments have been proposed for quick measurement of ^{15}N-labeled samples in solution (Lee, et al. 2006). This experiment is designed to suppress relaxation contributions from chemical exchange as well as dipole-dipole interactions with remote protons and the protein backbone amides. Data is collected as a series of one dimensional experiments, so data can be collected in a matter of minutes.

This method exploits the phenomenon of relaxation interference, the cross correlation between dipole-dipole and CSA relaxation mechanisms (Goldman, 1984), and is a simplified version of the relaxation interference experiments designed to identify exchanging residues discussed in the previous chapter. The ^{15}N nucleus in an amide generates two peaks corresponding to the α and β spin states, which have different

transverse relaxation rates. For a system with an axially symmetric ^{15}N chemical shift tensor, oriented with angle, θ, between its unique axis and the N-H bond, these rates can be written as:

$$R_\alpha = \lambda - \eta_{xy} + R_H + R_{ex}$$

and

$$R_\beta = \lambda + \eta_{xy} + R_H + R_{ex}$$

where λ is the auto-relaxation rate, η_{xy} is the transverse cross-correlated relaxation rate, R_{ex} is the exchange contribution to relaxation, and R_H is the contribution to relaxation from dipolar coupling with remote protons. The difference of the two rates, yields (Pervushin et al., 1997 and Lee et al., 2006):

$$R_\beta - R_\alpha = 2\eta_{xy} = 2p\delta_N(4J(0) + 3J(\omega_N))(3\cos^2\theta - 1)$$

where *p* describes the $^1H/^{15}N$ amide dipolar coupling:

$$p = \frac{\mu_0 \gamma_H \gamma_N h}{16\pi^2 \sqrt{2} r_{HN}^3}$$

where μ_0 is the permeativity of free space, γ_H and γ_N are the gyromagnetic ratios for ^1H and ^{15}N, respectively, *h* is Planck's constant, and r_{HN} is the ^1H-^{15}N internuclear distance (1.02 Angstroms). δ_N describes the CSA of the ^{15}N nucleus:

$$\delta_N = \frac{\gamma_N B_0 \Delta\delta_N}{3\sqrt{2}}$$

Where B_0 is the static magnetic field strength (16.4 T) and $\Delta\delta_N$ is the span of the axially symmetric ^{15}N chemical shift tensor (172 ppm). The spectral density function at frequency ω is:

$$J(\omega) = \frac{0.4\tau_c}{1+(\tau_c\omega)^2}$$

This spectral density is for a rigid spherical molecule, thus, this analysis may underestimate the tumbling time for flexible or non-spherical molecules.

The pulse sequence is written so that the transverse relaxation rates of the α and β states are collected separately. Each is measured at several delay times, and the relaxation rates decay exponentially with increasing delay time:

$$I_\alpha = e^{-R_\alpha \Delta}$$

and

$$I_\beta = e^{-R_\beta \Delta}$$

A simple fitting program allows extraction of R_α and R_β. Taking the difference of the rates allows straightforward determination of τ_c. The only trick is that the expression

for $2\eta_{xy}$ becomes quadratic in τ_c, with coefficients that render the $\tau_c\omega_N$ term negligible, so ω_N does not need to be known. Using p = -25.5 x 10^3 s^{-1} and δ_N = -1.8 x 104 s^{-1}, τ_c can be determined by solving the following equation:

$$0 = 4.49 \times 10^{26}\tau_c^2 - \eta_{xy}1.98 \times 10^{17}\tau_c + 3.97 \times 10^9$$

Because these are one-dimensional experiments, specific residues are not selected, rather, the area under the entire amide region (from about 6-11 ppm, depending on the protein) is integrated, so the results represent an average of all amides in the protein. It is also important to keep in mind that any residues that have broadened resonances due to intermediate exchange processes will not contribute to the overall result.

TRACT analyses were completed for RRM3, RRM4 and PTB1:34, and the results are given in Table 1. Data for RRM4 were collected using 1mM protein and gave a reasonable tumbling time of 6.4 ns. However, inspection of concentration dependent linewidths reveals broadening of RRM3 peaks at high concentrations. Therefore, all NMR data for RRM3 were collected using a 0.3mM sample. The tumbling time of 1.5 ns for RRM3 is clearly too low, and only proves that this method fails in the case of this protein.

No significant line broadening is apparent when 0.3mM and 1mM PTB1:34 are compared, however, an increase in the tumbling time determined by TRACT for the high concentration sample (10.8 ns for 1mM PTB1:34 as opposed to 7.0 ns for 300 μM

PTB1:34) suggests that some non-specific associations may be influencing the tumbling of the system. This concentration dependence of tumbling times for PTB1:34 is consistent with the results from ModelFree.

Protein-protein associations that influence the way the molecule tumbles in solution may give rise to artefactual exchange terms. Thus it is important to make sure that such associations are not responsible for R_{ex} terms that are interpreted as inherent molecular dynamics. There are a number of ways this may be done. NMR methods to resolve these issues have been discussed in the previous chapter. The simplest resolution exists if the molecule can be shown to tumble in solution as a monomer. Dynamic light scattering was used to test this, but inability of the data analysis program to fit the data in all cases except for the positive control (2μg/mL BSA, giving a hydrodynamic radius of 4.5) suggested that multiple species were in solution at all concentrations tested from 10 μM-1 mM for all proteins.

In the case where line broadening, light scattering and other techniques suggest that protein-protein interactions exist, collection of full relaxation data sets and ModelFree analysis of proteins at both high and low concentrations will reveal if such associations are responsible for predicted R_{ex} terms. Relaxation data for both RRM3 and PTB1:34 were collected at both 0.3mM and 1.0mM protein concentration (Figure 3). ModelFree analysis for RRM3 showed a decrease in tumbling time for the lower concentration (6.5 ns at 1mM and 5.1 ns at 0.3 mM). At the high concentration, almost all residues required R_{ex} (thought most were very small, ie:less than 3 s^{-1}), whereas at the

low concentration, only two residues required R_{ex} terms. This is in contrast to PTB1:34, where tumbling time was decreased at the low concentration but R_{ex} terms were required throughout the protein at both concentrations.

	Concentration	τ_{MSE}	$\tau_{MModelFree}$	τ_{MTRACT}	$\tau_{MHYDRONMR}$
RRM3	0.3mM	5.2ns	5.1ns	1.5ns	7.5ns
RRM4	1.0mM	4.5ns	6.0ns	6.4ns	6.1ns
PTB1:34	0.3mM	9.7ns	7.2ns	7.0ns	32.3ns
PTB1:34	1.0mM	9.7ns	9.7ns	10.8ns	32.3ns

Table 1: Summary of tumbling times for PTB1:34, RRM3 and RRM4. Stokes-Einstein calculations, ModelFree fits, TRACT experiments and HYDRONMR calculations were used to evaluate the constructs. RRM3 experiments were performed at 0.3mM since higher concentrations were known to cause NMR linebroadening and artefactual R_{ex} terms in the ModelFree fit. Data for RRM4 were collected using 1mM protein as no linebroadening effects and few R_{ex} terms were predicted by ModelFree. Data for PTB1:34 were collected at both high and low concentrations to ensure that the concentration dependence of the tumbling times was not the cause of the observed pervasive R_{ex} terms.

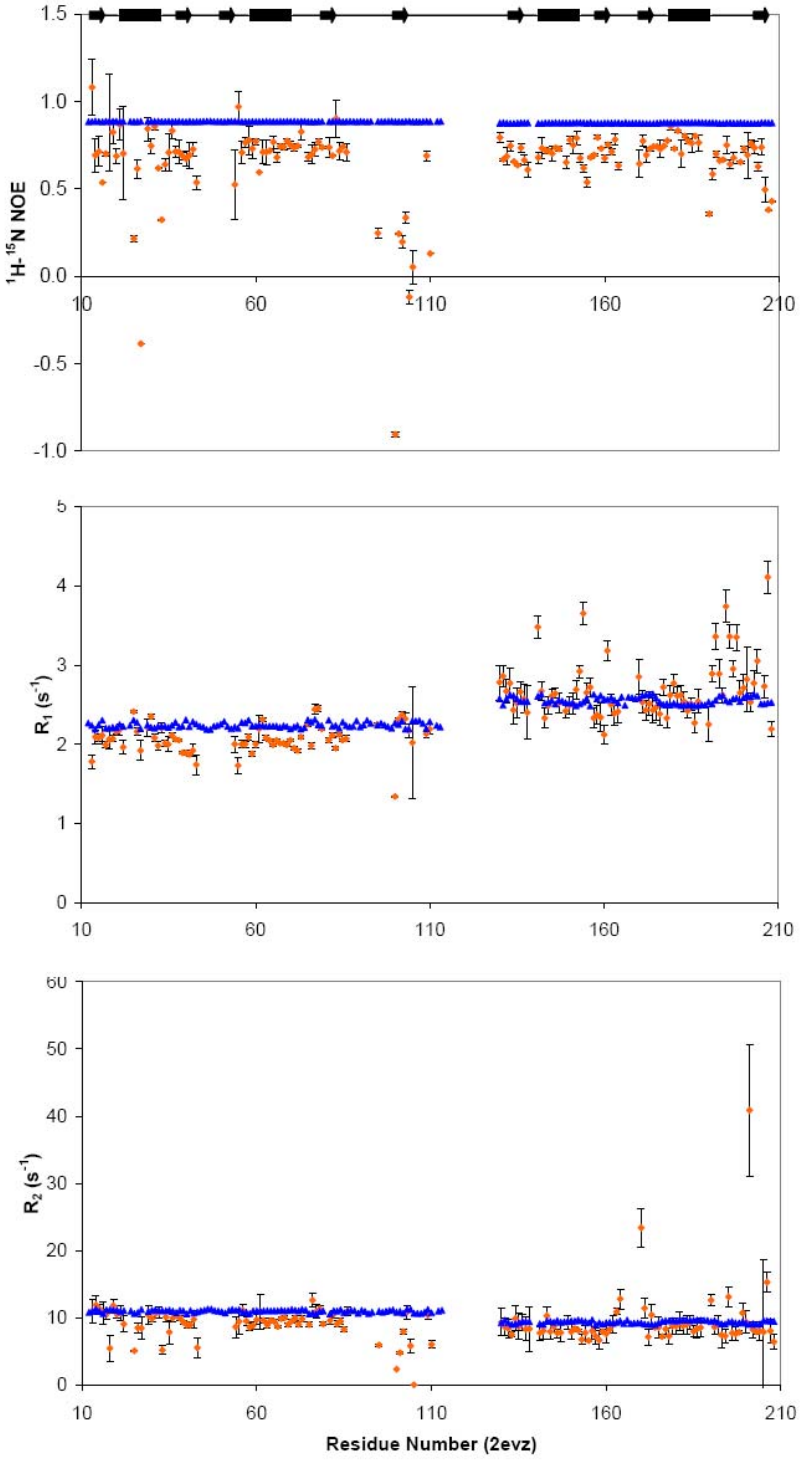

1.5
1.0
0.5
0.0
-0.5
-1.0
^{1}H-^{15}N NOE
10
60
110
160
210
5
4
3
2
1
0
R_1 (s^{-1})
60
50
40
30
20
10
0
R_2 (s^{-1})
Residue Number (2evz)

Figure 1: HYDRONMR predicts NMR relaxation values reasonably well for the individual domains. Comparison of RRM3 and RRM4 NMR R_1, R_2 and ^{1}H-^{15}N-NOE data from experiment (orange diamonds) and hydronmr (blue triangles). Numbering is from PDB ID:2EVZ, and secondary structure elements are shown at the top with beta strands as arrows and alpha helices as cylinders.

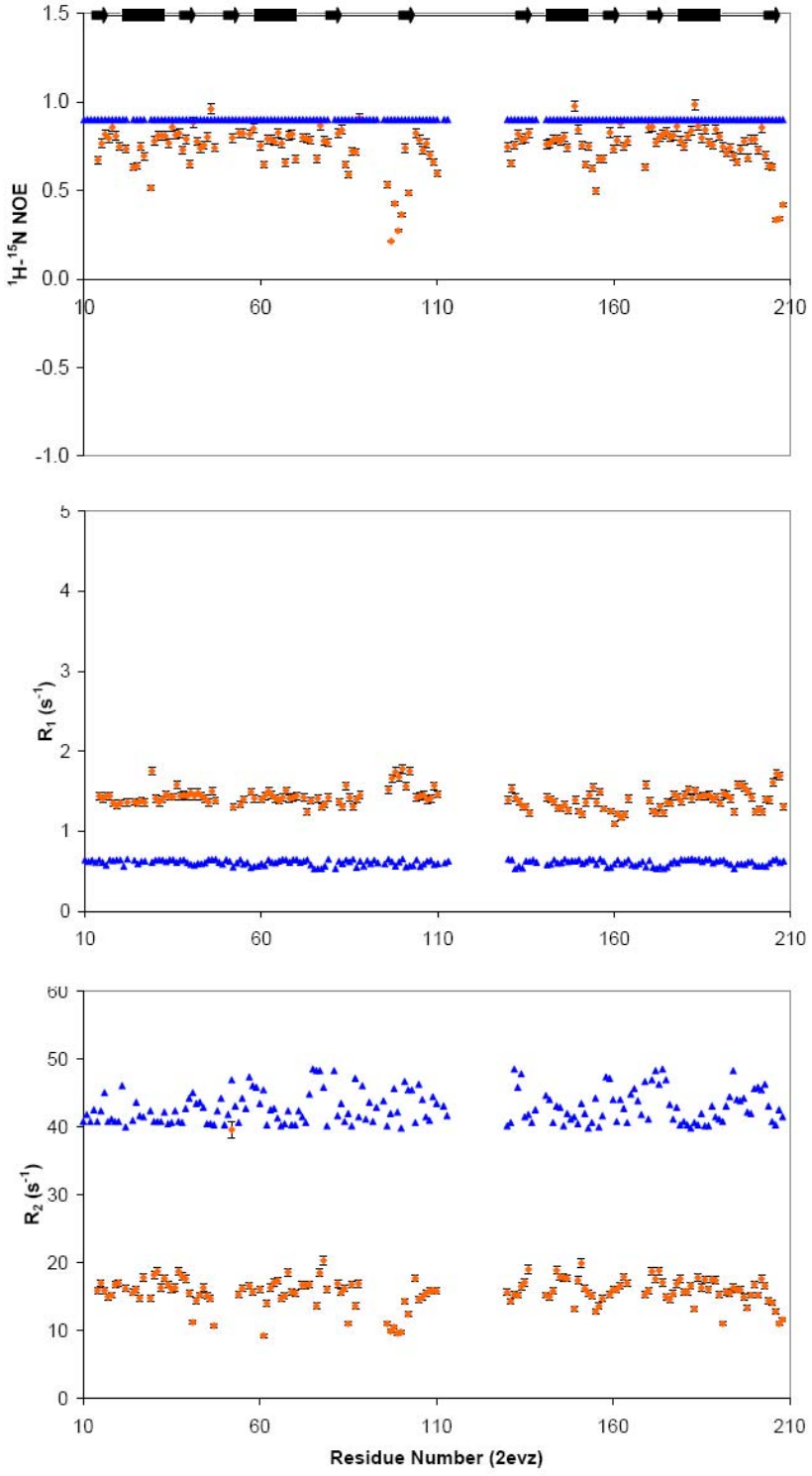

1.5
1.0
0.5
0.0
-0.5
-1.0
^{1}H-^{15}N NOE
10
60
110
160
210
5
4
3
2
1
0
R_1 (s^{-1})
60
50
40
30
20
10
0
R_2 (s^{-1})
Residue Number (2evz)

Figure 2: HYDRONMR completely fails to predict accurate NMR relaxation values for PTB1:34. Comparison of PTB1:34 NMR R_1, R_2 and ^{1}H-^{15}N-NOE data from experiment (orange diamonds) and hydronmr (blue triangles). Numbering is from PDB ID:2EVZ, and secondary structure elements are shown at the top with beta strands as arrows and alpha helices as cylinders.

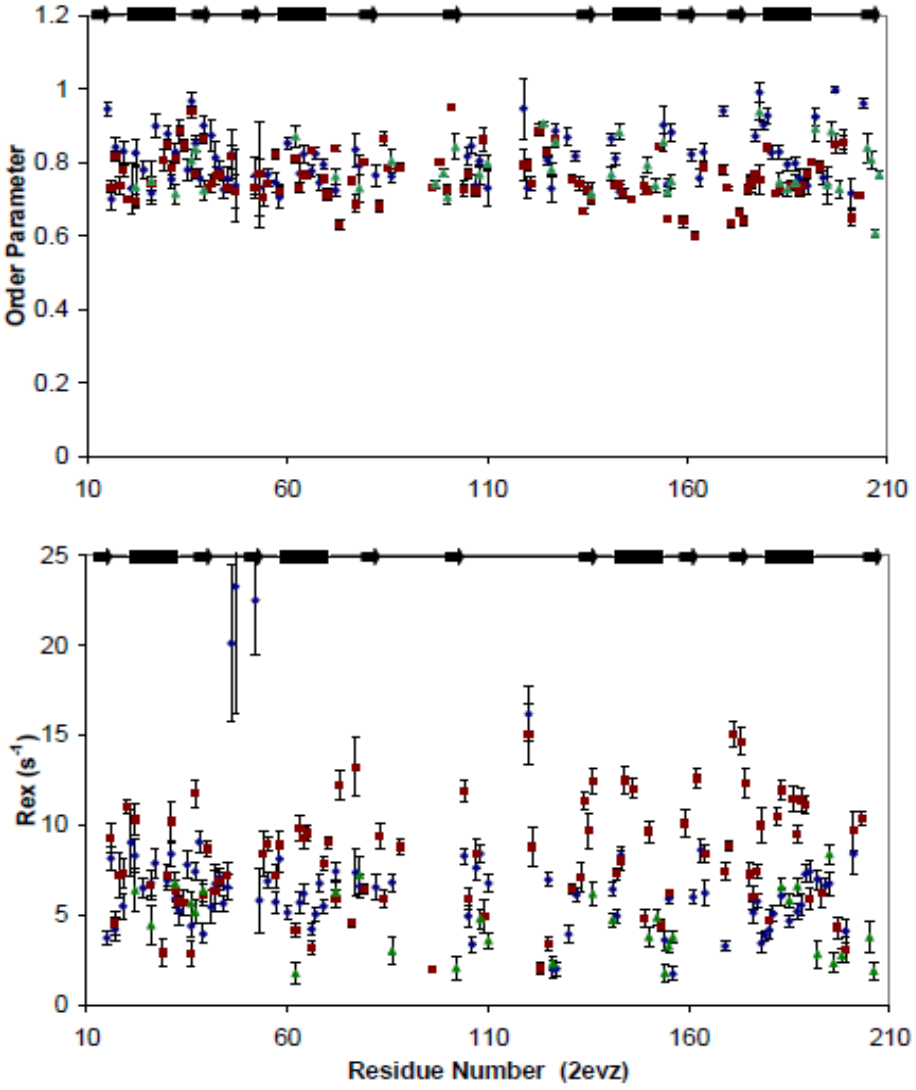

Order Parameter
1.2
1
0.8
0.6
0.4
0.2
0
10
60
110
160
210
Rex (s-1)
25
20
15
10
5
0
10
60
110
160
210
Residue Number (2evz)

Figure 3: Concentration effects do not explain the presence of pervasive R_{ex} terms in PTB1:34. Order parameters (top panel) and R_{ex} terms (bottom panel) from ModelFree fits of 300μM PTB1:34 at 700MHz (green triangles), 1mM PTB1:34 at 700MHz (blue diamonds), and 1mM PTB1:34 at 500 and 600MHz (red squares) are plotted against residue number (PDB ID: 2EVZ) with secondary structure elements indicated at the top. While there is little variation, within error, for order parameters and R_{ex} terms between the two data sets fit at 700MHz, the difference in tumbling times reflects a degree of protein self-association with τ_M= 7.2 ± 0.06 ns at 300μM and τ_M = 9.2 ± 0.10 at 1mM. There are much more significant differences between the fit at two static magnetic field strengths verses the fits at a single static magnetic field strength. Importantly, these data show that while under-fitting (ie: determining R_{ex} from data at a single static magnetic field strength) appears to be a problem, it is not the source of the pervasive R_{ex} terms reported. Furthermore, differences between the data sets at 300μM and 1mM at 700MHz, are localized to residues flanking the loops, suggesting that, while protein self-association may be the source of some exchange terms, it does not explain the bulk of the slow motions observed in PTB1:34.

References:

Garcia de la Torre, J., Huertas, M. L., and Carrasco, B. HYDRONMR: Prediction of NMR relaxation of globular proteins from atomic-level structures and hydrodynamic calculations. (2000) *J. Mag. Res.*, **147**, 138-46.

Goldman, M. Interference effects in the relaxation of a pair of unlike spin-1/2 nuclei. (1984) *J. Magn. Reson.*, **60**, 437-452.

Lee, D., Hilty, C., Wider, G., & Wüthrich, K. Effective rotational correlation times of proteins from NMR relaxation interference. (2006) *J. Magn. Reson.*, **178**, 72-76.

Pervushin, K., Riek, R., Wider, G. & Wüthrich, K. Attenuated T2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. (1997) *Proc. Natl. Acad. Sci.*, **94**, 12366-12371.

Ryabov, Y. E.,Geraghty, C., Varshney, A., Fushman, D. An efficient computational methods for predicting rotational diffusion tensors of globular proteins using an ellipsoid representation. (2006) *J. Am. Chem. Soc.* **128**, 15432-44.

Vitali, F., Henning, A., Oberstrass, F. C., Hargous, Y., Auweter, S. D., Erat, M., Allain, F. H-T. Structure of the two most C-terminal RNA recognition motifs of PTB using segmental isotope labeling. (2006) *EMBO J.*, **25**, 150-62.

Chapter 5

Interactions between PTB RRMs induce slow motions and increase RNA binding affinity

Maynard, C. M. and Hall, K. B.

J. Mol. Biol., 397, 260-277.

Abstract

Polypyrimidine tract binding protein (PTB) participates in a variety of functions in eukaryotic cells, including alternative splicing, mRNA stabilization, and internal ribosomal entry site (IRES) mediated translation initiation. Its mechanism of RNA recognition is determined in part by the novel geometry of its two C-terminal RNA Recognition Motifs (RRM3 and RRM4), which interact with each other to form a stable complex (PTB1:34). This complex itself is unusual among RRMs, suggesting that it performs a specific function for the protein. In order to understand the advantage it provides to PTB, the fundamental properties of PTB1:34 are examined here as a comparative study of the complex and its two constituent RRMs. Both RRM3 and RRM4 adopt folded structures, but RRM3 readily self-associates. The RNA binding properties of the domains differ dramatically. RRM4 does not bind to RNA, and although RRM3 binds to polypyrimidine tracts, its affinity is significantly weaker than that of PTB1:34. ^{15}N-NMR relaxation experiments show that PTB1:34 has slow, microsecond motions throughout both RRMs including the interdomain linker. This is in contrast to the individual domains, RRM3 and RRM4, where only a few backbone amides are flexible on this timescale. The slow backbone dynamics of PTB1:34, induced by packing of RRM3 and RRM4, could be essential for high affinity binding to a flexible polypyrimidine tract RNA and also provide entropic compensation for its own formation.

Introduction

In recent years there have been a significant number of studies that relate molecular motions to the functions of biological molecules. Examples include such processes as enzyme catalysis, protein-ligand interactions, and both inter- and intra-molecular interactions between proteins[1-8]. Motions on a biologically relevant time-scale can vary from picoseconds to seconds, and occur both proximal and distal from a given interaction site. NMR relaxation methods are able to probe residue specific motions across this wide range of time scales, and thus are ideal techniques for gleaning detailed information about the importance of motions of biological molecules. The continuing challenge is to understand which molecular motions are functionally relevant, and for that assessment, there must be a means to compare and contrast motions with function.

RNA Recognition Motifs (RRMs, also known as RNA Binding Domains or RBDs) provide an example of how backbone dynamics and function can be linked. The RRM is the most common eukaryotic RNA binding domain, with over 150 structures deposited in the PDB to date. A typical RRM has βαββαβ secondary structure folded into the α/β sandwich tertiary fold, with a four-stranded antiparallel β-sheet. Very little variation in three dimensional structure is seen across the family[9]. In addition, RRMs are characterized by conserved RNP1 and RNP2 sequences that contain several aromatic side chains displayed on the surface of the β-sheet (Fig. 1). In general, these amino acids stack with RNA bases during binding, and thus define the canonical RNA binding surface. The conservation of sequence and three dimensional structure leads to the obvious question of how an RRM selects a specific RNA target.

The most detailed studies of RRM binding have focused on the human U1A protein, which specifically binds with high affinity to an unstructured seven nucleotide RNA sequence. The U1A RRM has been shown to undergo rapid correlated motions that organize its RNA binding surface[10], and mutations that perturb its dynamics have been shown to weaken its affinity for RNA[11-13]. The role of backbone dynamics in the RNA binding of other RRMs has not been investigated, so its generality remains to be established. However, for RRMs that recognize simple RNA sequences such as poly(A) binding protein[14] or U2AF[15] that also rely on several tandem RRMs to confer affinity and specificity, backbone dynamics could have little functional contribution to RNA binding.

The RNA sequences bound by Polypyrimidine tract binding protein (PTB) are simple: typically uridine-rich, with interspersed cytosines. The lengths of these sequences can vary enormously, yet PTB is able to bind to (U/C) tracts from tetramers to hundreds of nucleotides. PTB contains four RRMs, all of which differ significantly from canonical RRMs in terms of their RNP sequences [16,17] (Fig. 1). Specifically, PTB RRMs have hydrophobic side chains replacing the solvent accessible aromatic amino acids on the β-sheet surface. At the junction of β3 and loop 3, a highly conserved glycine, thought to act as a hinge for the loop, is replaced with a much larger amino acid[18]. The tertiary structures of RRM2 and RRM3 differ from the canonical RRM as well, for both have a fifth β-strand that packs against β2 via a long loop that spans the β-sheet surface connecting β5 to β4. This additional strand extends the canonical RNA binding surface [19, 20] at the same time as the connecting loop occludes it, presenting rather a conundrum regarding β-sheet function. These noncanonical RRMs are phylogenetically conserved in

PTB proteins, suggesting a novel mode of RNA recognition.

The four RRMs of PTB are not equivalent in their contributions to RNA binding, or in their relative geometry in the protein. RRM1 and RRM2 are separated by a 25 amino acid linker, and are each able to bind RNA as independent domains [19, 21]. RRM2 and RRM3 are separated by an 80 – 100 amino acid linker, effectively separating the two N-terminal RRMs from the two C-terminal RRMs. The variation in this middle linker results from alternative splicing that produces three PTB isoforms: PTB1, PTB2, and PTB4, which have conserved RRMs but insertions in the RRM2-RRM3 linker. The C-terminal RRM3 and RRM4 are separated by a 24 amino acid linker, but in this case, the two RRMs interact extensively [20, 22] and their linker is an intrinsic part of their structure. The interface between RRM3 and RRM4 involves both helices of RRM3, one helix and β4 of RRM4, and the linker. The orientation of RRM3 and RRM4 places their β-sheet surfaces in opposing directions and therefore imposes a length constraint on a single RNA strand that would bind to both surfaces.

Among characterized proteins with multiple RRMs, only hnRNPA1, Prp24, and PTB have been shown to exhibit RRM:RRM interactions that lead to stable intramolecular complexes [20, 22-25]. PTB RRM3 and RRM4 together form a stable domain that constitutes the free and bound forms of the protein. RNAs preferentially bound by the two C-terminal RRMs (PTB1:34) contain unstructured $(U/C)_n$ tracts from n = 11 ($GABA_A$ γ2 intron) to n = 120 (HCV 3' NTR)[26]. Short (U/C) tracts separated by poly$(rA)_n$ spacers of variable length were found to bind with highest affinity to PTB1:34

with a spacer of n=15 [22]. The variation in RNA targets implies a plastic binding surface of PB1:34 that would be necessary to accommodate their sequences, lengths, and structures, although the details of its interaction with long polypyrimidine tracts are not known.

To understand the intrinsic properties of PTB1:34, we compare it to its two RRM constituents using NMR and ^{15}N-NMR relaxation experiments. RNA binding properties of PTB RRM3 and RRM4 and PTB1:34 are compared using one of the known pre-mRNA targets of PTB. All data show that PTB1:34 is far more than the sum of its parts, for it binds with high affinity to the RNA, a property that we propose is due in part to a complete reconfiguration of the backbone dynamics to create a novel binding platform for RNA.

Results

Although the two C-terminal RRMs of human PTB interact with each other through a stable interface, they can be studied as separate RRMs. Constructs of each RRM including a section of the intervening linker were prepared such that when the RRMs were mixed, all but two residues of the entire linker sequence were present.

RNA Binding. The relative RNA binding affinities of PTB1:34 and the individual RRMs were compared using a 120 nucleotide RNA from the rat $GABA_A$ γ2 pre-mRNA, a natural target of PTB[27], which has previously been shown to bind PTB1:34 with high affinity [28]. This RNA is predicted to be single-stranded with no stable secondary structure, so that its two polypyrimidine tracts are accessible to the protein. Footprinting

experiments showed that PTB1:34 protects both the 34-nt tract and the 11-nt tract of the RNA, even at 10 nM protein[28].

The stoichiometry of PTB binding to this RNA is not 1:1, so Electrophoretic Mobility Shift Assay (EMSA) was used to assess the relative binding affinities. As shown in Figure 2, PTB1:34 binds with high affinity to the RNA (binding is observed here at 10 nM protein), while RRM3 binding is approximately 50-fold weaker. The stoichiometry of RRM3 binding has not been determined, but as the gels show, several complexes are observed at higher concentrations of protein. Whether these higher order complexes result from protein:RNA or protein:protein interactions is not known. In contrast to RRM3, RRM4 does not bind to this RNA under any conditions tested (50-200 mM NaCl) at concentrations up to 5 μM. In an attempt to restore affinity of RRM3, equimolar RRM3 and RRM4 constructs were mixed and bound to the RNA. Those data (Figure 2) show that the presence of RRM4 neither restores nor reduces the affinity of RRM3 for this RNA, indicating that PTB1:34 has unique RNA binding ability.

The properties of PTB1:34 that allow it to bind with high affinity to this RNA could include a unique electrostatic potential surface that attracts the RNA strand. Certainly one feature of a polypyrimidine tract with a high proportion of uridines is its flexibility, and it could wrap around PTB1:34 to make contacts between its phosphate backbone and positively charged amino acid sidechains. Indeed, RNA binding is salt dependent over the range of 50 to 500 mM NaCl (although affinities have not been quantified) indicating that some electrostatic component does contribute to complex

formation.

To characterize the surface electrostatics of PTB1:34, RRM3, and RRM4, each construct was analyzed using the adaptive Poisson-Boltzman solver (APBS) to calculate the electrostatic potential surface[29]. The results were mapped onto the solvent accessible surface area of each protein (Fig. 3), and show that the interaction between RRM3 and RRM4 creates a distribution of positive potential on PTB1:34 that could be a path for the RNA backbone, and may contribute to RNA binding. This path includes amino acids in the linker, and although direct interactions between RNA and linker residues have not been reported, they cannot be excluded especially when the RNA strand is long enough to wrap around the domain. Charge polarization of the PTB1:34 domain localizes a negative potential on one side of RRM4 and positive surfaces on RRM3 and RRM4 that could be important for organizing and arranging protein:protein interactions in multimeric complexes[28].

Protein structure and stability

The structure of PTB1:34 has been solved by NMR[19, 21-23], but there are no structures of the individual RRMs. To compare the secondary structures of the proteins, CD spectra were measured (Figure S1). Each RRM has a stable secondary structure, suggesting that each has adopted the predicted tertiary fold. Two features of the constructs are worth noting, however. The first is that the spectrum of mixed RRM3 + RRM4 does not reproduce the spectrum of PTB1:34. In particular, the molar ellipticity per residue of PTB1:34 is lower than that of either RRM alone. The tertiary structure of PTB1:34 is

rather notable for its low fraction of folded protein; only 35% of the residues are part of β-strand or α-helix[22]. A large fraction of PTB1:34 residues is found in the interdomain linker and loops.

The thermodynamic stability of the proteins was compared using chemical denaturation and thermal melting. Guanidine hydrochloride denaturation profiles of all protein constructs were measured by monitoring the CD signal at 222 nm (Figure S1). RRM3 denaturation can be fit by a two-state model, with an unfolding free energy of -6.3 ± 0.6 kcal/mol. However, RRM4 and PTB1:34 denaturation curves are more complex. In particular, the denaturation profile of RRM4 did not show a clear transition, but was non-cooperative from 0 to 7 M GndHCl. This type of denaturation profile has been reported for proteins that are "downhill folders"[30]. Such proteins are characterized by a broad landscape of incremental free energy wells that could allow the structure to be adaptable over a range of environments. The denaturation curve of PTB1:34 likewise cannot be fit by a two-state transition. We speculate that the initial increase in negative ellipticity could arise from separation of the two RRMs within PTB1:34, which then denature with their characteristic profiles. Thermal denaturation of the proteins did not allow additional thermodynamic characterization, since RRM3 and PTB1:34 thermal melts are not reversible at micromolar concentrations. RRM4 is not thermally denatured at 90 °C at pH 6.8, indicating again that RRM4 has distinctive properties that could be critical for formation of PTB1:34.

The structure of PTB1:34 has been solved by NMR[19-22], but under sufficiently

different solution conditions that the NMR backbone assignments had to be repeated here. In experimental conditions used for RNA binding (20 mM potassium phosphate, pH 6.8, and 100 mM KCl), many backbone amide protons in the interdomain linker, the loop between β2 and β3 of both RRMs, and the loop between RRM3 β4 and β5 were not observable on the chemical shift time scale. A comparison of the published chemical shifts[19] and our assignments for PTB1:34 indicates that the resonances are sensitive to solution conditions, but for subsequent experiments, the structure was assumed to be as reported.

For our experiments, the backbone $^{1}H/^{15}N$ resonances of RRM3 and RRM4 were assigned using standard NMR methods, but full structure determinations have not been done. However, the assigned portions of the constructs indicate that the two RRMs adopt similar structures alone and in PTB1:34, so the NMR structure[20] of PTB1:34 is used as a template for further comparisons. The structure of the interdomain linker obviously differs in the two RRMs and in the PTB1:34 domain; in the two RRMs it is disordered but becomes more ordered in PTB1:34.

The $^{1}H/^{15}N$-HSQC spectra of the protein constructs reveal several important features of their structure and stability (Figure S2). The $^{1}H/^{15}N$-HSQC spectrum of RRM4 is consistent with that of a folded protein. Most resonances are assigned, with the exception of the β2-β3 loop. Notably, most resonances in the $^{1}H/^{15}N$-HSQC spectrum of RRM4 are readily identified in the spectrum of PTB1:34, many of them being superimposable. The structure of RRM4 has clearly not been significantly perturbed in

the context of PTB1:34.

The $^{1}H/^{15}N$-HSQC spectrum of RRM3 is complex, and while the proton chemical shift dispersion is consistent with that of a folded protein, several residues have more than one resonance. These backbone amides are in slow exchange on the NMR chemical shift timescale, which indicates conformational heterogeneity of this RRM. The structural heterogeneity persists from 4 to 37 °C in 100 mM KCl, 20 mM sodium phosphate pH 6.8. It is worth noting that the conformational sampling is not apparent in its two-state unfolding curve (Figure S1). Approximately 70% of RRM3 amides are assigned; most missing residues are those in and around loop3 between β2 and β3, and the loop that connects β4 with β5, where amides are in exchange with solvent and absent from the spectrum. In the context of PTB1:34, RRM3 loses its conformational heterogeneity and shows single amide resonances. It is not surprising that most of its backbone amide resonances are not superimposable with their PTB1:34 counterparts.

A comparison of the assigned resonances of RRM3 and RRM4 with those of PTB1:34 shows that the major chemical shift changes are within the extensive interface (Figure 4). Since the interface involves both helices of RRM3, those amide chemical shifts are expected to change due to their new environment, and as Figure 4 illustrates, there are also chemical shift changes in the amides of the one helix and β4 of RRM4. For RRM4 most chemical shift changes are minor, but for RRM3, more changes are significant, undoubtedly arising from stabilization of the RRM3 structure. We conclude that the RRMs free and in the PTB1:34 protein have the same global folds.

The PTB1:34 interface is composed of both RRMs and the linker. The contribution of the linker to the physical connection of RRMs is clear from NMR experiments that mix the two RRMs. At concentrations up to 0.5 mM of each RRM, PTB1:34 was not spontaneously formed when the individual domains were mixed. In these solution conditions of 100 mM KCl, 20 mM sodium phosphate, pH 6.8, there was no evidence of complex formation at temperatures from 10-40 ˚C, and after eight months of incubation at room temperature. Conversely, heating PTB1:34 to 45 ˚C does not separate the two RRMs. The role of the 24 amino acid linker in stabilizing and orienting the two RRMs is clearly significant.

Protein backbone dynamics

Formation of the intricate interface between the two RRMs obviously led to the elimination of the conformational heterogeneity of RRM3 observed in the $^{1}H/^{15}N$-HSQC, but could also have resulted in less apparent changes in the dynamics of the two RRMs. Since the dynamics of U1A RRMs have been implicated in their RNA binding mechanism[11-13, 22, 31, 32], the PTB constructs were compared to determine if their backbone dynamics differed.

Fast motions and global tumbling

Standard R_1, R_2 and $^{1}H/^{15}N$ heteronuclear NOE experiments[33] were used to measure fast (ps-ns) backbone dynamics of RRM3, RRM4 and PTB1:34 at 25˚C (Figure S3). The measured R_2/R_1 ratios were used to calculate the rotational correlation times of each molecule using ModelFree. Individual domains were best fit by an isotropic

diffusion tensor to yield tumbling times (τ_m) of 5.12 ± 0.05 ns for RRM3 and 5.98 ± 0.02 ns for RRM4. Each RRM retains a portion of the interdomain linker, which probably increases their global tumbling times, but fitting to an axially symmetric model did not converge. PTB1:34 was best fit with an axially symmetric tensor ($D_{\parallel}/D_{\perp} = 1.6 \pm 0.2$) to give $\tau_M = 9.6 \pm 0.1$ ns.

The overall correlation time (τ_m) is a critical parameter for fitting the relaxation data, as is the description of the diffusion tensor. To supplement the calculations from ModelFree, HYDRONMR[34] calculations and TRACT[35] experiments were used for all three proteins. Although neither of these analyses was able to successfully determine tumbling times to compare to the ModelFree fits, they do report on the consistency of the underlying assumptions of domain rigidity and the concentration dependence of the tumbling time.

HydroNMR is a software package which estimates the tumbling time of a molecule in solution based on beads packed around a rigid structure[34]. We used the smallest minibeads to model the domains: bead diameters of 1-2 Å were used for RRM3 and RRM4 and 1.5-3.0 Å for PTB1:34. This analysis yields a tumbling time of 6.1 ns for RRM4, in good agreement with the ModelFree fit. The tumbling time estimated for RRM3 was 7.5 ns, which is slightly more than the tumbling time extracted from ModelFree, but does make sense when the differences in the methods (i.e.: peripheral flexible loops in RRM3 would cause an overestimation of tumbling time for HydroNMR), as well as the slightly higher molecular weight of RRM3 is taken into

account. This method roundly fails when applied to PTB1:34, giving a tumbling time of 32.3 ns. Clearly this is a nonsensical result, and indicates that PTB1:34 is a more flexible molecule than RRM3 or RRM4. HydroNMR also is able to back-calculate R_1, R_2 and heteronuclear NOE values. Here again, it is evident that the HydroNMR estimates agree reasonably well with experimental values for the individual RRMs, but not for PTB1:34 (data not shown).

Experimentally, NMR TRACT experiments can be used to measure overall correlation times[35]. This method also assumes a rigid molecule, but gives a lower bound. Tumbling times obtained from this method give 6.4 ns for 1 mM RRM4, again in good agreement with other methods and supporting the picture of RRM4 as a packed globular domain. However, tumbling times of 1.5 ns and 7.0 ns for 300 μM RRM3 and 1 mM PTB1:34, respectively, were much lower than expected, indicating that for these constructs the rigid molecule assumption fails.

As another metric of the assumption of noninteracting domains in the NMR samples, inspection of NMR linewidths in $^{1}H/^{15}N$-HSQC experiments for protein concentrations from 100 μM to 1 mM shows that only RRM3 has a concentration-dependence. This construct appears to self-associate at the lowest concentrations measurable, as demonstrated by the average R_2/R_1 ratio for RRM3, which at 1 mM is approximately double that of the ratio at 300 μM. RRM3 self-association must be considered in analysis of all relaxation measurements, even though data were collected at 300 μM protein.

A comparison of linewidths for RRM4 and PTB1:34 failed to show a concentration dependence, suggesting that these two proteins behave as monomers in solution. Using dynamic light scattering (DLS) to assess the homogeneity/heterogeneity of the proteins in solution strongly indicates that PTB1:34 is polydisperse in solution at concentrations from 10 μM to 1 mM. Its scattering profile cannot be fit to a single species, and although we could not fit the data to multiple species to estimate their populations, it is clear that this protein has a propensity to self-associate. It is worth noting that PTB1:34 solution structure was solved in conditions of 20 mM phosphate buffer, pH 6.8 and 100 mM KCl at a concentration of 1 mM. As expected, RRM3 also shows a complex DLS profile consistent with self-association. Rather surprisingly, RRM4 profiles also indicate the presence of larger species, even though its solution properties are otherwise those of a single species. Unfortunately, DLS data cannot be interpreted in terms of a proportion of different species in solution. The control BSA sample consistently gave a scattering profile of a monomer, so we must conclude that all constructs have a propensity to self-associate in this buffer.

Slow backbone motions

To describe the domains, the NMR data were fit using the Lipari-Szabo formalism[36,37] to give the order parameter, S^2, and the exchange term, R_{ex}. These parameters describe the local reorientation and slow motions, respectively, of each amide N-H vector. Order parameters range from $0 \leq S^2 \leq 1$, with $S^2 = 0$ indicating isotropic free rotation and $S^2 = 1$ indicating a rigid vector within the molecular frame. Whereas order parameters report on fast motions (ps-ns), exchange terms indicate μs-ms (slow) motions.

Here we use S^2 values to compare backbone amide mobility and R_{ex} terms to indicate the presence of slow motions.

Fits of the relaxation data (Figure 5) make it clear that each protein has unique backbone dynamics. Of the two RRMs, the body of RRM4 has uniformly high order parameters ($S^2_{av} = 0.85$), indicating an overall rigid backbone. Only residues flanking the loop between β2 and β3 require R_{ex} terms for fitting, indicating the presence of slower motions (μs – ms). Order parameters for RRM3 also indicate a rigid protein, however, several residues have order parameters greater than the theoretical limit (0.95), indicating that the fit is unreliable. In addition to the residues that are not assignable under these conditions, several residues were not able to be fit at all by ModelFree. Given the conformational fluctuations of RRM3 that are apparent in its amide $^{1}H/^{15}N$-HSQC spectrum, and the fact that this construct appears to self-associate, even at concentrations as low as 10 μM, only a subset (about 40%) of all amide resonances yield reliable data for this protein construct.

The same analysis of PTB1:34 shows an overall change in the dynamics, for here both RRMs have greatly increased backbone flexibility. PTB1:34 order parameters are lower, and most of its amides require exchange terms to describe their motions. Modelfree analysis of the relaxation data indicates that slow motions are uniformly distributed throughout PTB1:34, including both RRMs and their interdomain linker. This is an unexpected result and required more extensive experiments for verification.

Pervasive R_{ex} terms can arise as an artifact of an inappropriate diffusion tensor or from global motions of the molecule (such as flexing about the interface). In PTB1:34, we suspect that the diffusion tensor is time-dependent, given the large proportion of long loops and tails and the small fraction of residues that comprise stable secondary structures. The uncertainty in the appropriate description of the diffusion tensor led to direct measurements of the exchange contribution to transverse relaxation using standard NMR relaxation experiments.

Relaxation interference experiments[38] were used to determine the exchange free transverse relaxation rate of PTB1:34 (Figure 6). Data were collected at the highest static magnetic field strength available (700 MHz) in order to maximize the CSA effect. The scaling factor, κ, which takes into account autorelaxation effects, was calculated as the theoretical ratio of R_2/η_{xy} for $^1H/^{15}N$ dipolar and ^{15}N CSA relaxation pathways. Assuming that the dipolar and CSA principle axes are co-linear, and that the molecule is large enough so that only J(0) needs to be considered, κ=1.06 at 700 MHz. These experiments do not allow characterization of the timescales of motions, but the data do provide a robust identification of residues which undergo exchange rates from μs-ms, thus confirming the presence of slow motions throughout the protein. A similar approach to characterizing motions was taken by Pervushin et al.[39], where they were able to show pervasive motions throughout a molten globule-like protein. PTB1:34 is not a molten globule, but its core tertiary fold is not large in proportion to its loops which have complex motions on many timescales. The success of the analysis was apparent from recalculations of ModelFree analysis of PTB1:34 using the exchange free transverse

relaxation rate, $\kappa\eta_{xy}$, in place of the standard transverse relaxation rate, (R_2). Without the bias from the R_{ex} terms, the calculations returned similar order parameters and overall tumbling time (τ_m), but only a single residue in RRM4 that suggested motion on a slow timescale. These data further support the conclusion that the R_{ex} terms from fits to the data are reporting on slow global motions and are not an artifact of an inappropriate diffusion tensor.

Relaxation interference experiments were also performed for 1 mM RRM4 at 700 MHz. Using κ=1.35 obtained from the trimmed mean[40] of the ratio of R_2/η_{xy} (this domain has a relatively large and stable core structure), only seven residues with R_{ex} greater than 5 Hz (data not shown) were reported. This result is consistent with the ModelFree predictions.

The R_{ex} term from the Lipari-Szabo formalism cannot provide precise values of the exchange time; it only indicates that motions on the μs-ms timescales are required to fit the data. To measure the timescale of the slow motions in PTB1:34, NMR ^{15}N-CPMG (to assess millisecond motions), and ^{15}N-$R_{1\rho}$ (to assess microsecond motions) data were collected at 700 MHz to maximize exchange contribution to the transverse relaxation rate. Analysis of CPMG experiments indicated that there were no millisecond motions in the assignable regions of the protein at 25 °C (data not shown). Preliminary ^{15}N-$R_{1\rho}$ data indicate that microsecond motions are present throughout the protein, in qualitative agreement with the ModelFree calculations and the relaxation interference experiments.

^{15}N-$R_{1\rho}$ relaxation experiments can be used to directly measure μs motions, and are not contingent on a description of the molecular tumbling time. To ensure that the motions detected in the relaxation interference experiments and the ModelFree fits are not artifacts of data analysis, transverse relaxation in the rotating frame was used to directly detect microsecond exchange. $R_{1\rho}$ relaxation rates depend on spin-lock field strengths in the presence of microsecond exchange, and a dispersion curve can be used to extract information about populations, exchange rates and chemical shift differences for a two state system. However, this system is not likely to be two state, and is thus not amenable to fitting to this type of curve. Here, we define $\Delta R_{1\rho}$, the difference between $R_{1\rho}$ relaxation rates at two different spin lock field strengths (Figure 7). This method is not able to extract quantitative information about the states, and is prone to underestimating the exchange contribution to the transverse relaxation rate since a fit cannot be used to extrapolate to a spin lock field strength of 0 Hz, but does robustly identify residues which experience exchange on a microsecond timescale. $\Delta R_{1\rho}$ was determined by measuring transverse relaxation in the rotating frame, on resonance, at two field strengths, 350 Hz and 1750 Hz. Residues throughout PTB1:34 were found to experience microsecond exchange, and is in qualitative agreement with the R_{ex} terms from the ModelFree fits and relaxation interference experiments.

This reorganization of protein backbone dynamics of PTB RRM3 and RRM4 upon formation of PTB1:34 is striking and indicates that the motions are an important property of the functional complex. A visual comparison of the backbone dynamics of the RRMs in their free and complex states illustrates the extent of the changes (Figure 8).

Although the interaction between RRM3 and RRM4 may contribute to protein function via modest structural reorganization, the biological implications of the motional reorganization are of particular interest. We propose that the microsecond dynamics of the PTB1:34 backbone have two functions: one to provide entropic compensation for the protein:protein association, and the other to prime the protein to select a very flexible RNA as a binding target.

Discussion

PTB has been implicated in such a wide variety of biological functions[41] that understanding its RNA selection mechanism is critical for predicting its role in a specific environment. Here, we focus on only the two C-terminal RRMs of PTB, yet this half of the protein has the capacity to act independently of its N-terminal domains due to the long, flexible linker that connects RRM3 to RRM2[20, 42]. Indeed, there are two reports that suggest a truly independent function for PTB1:34. The first showed that the polio protease $3C^{pro}$ can cleave the linker between RRM2 and RRM3[43]. Since PTB has been shown to be required for translation initiation at the poliovirus IRES[44], this cleavage event is intriguing in its implication of a separation of function. The second report identified a new alternative splicing event that produces an independent PTB1:34 protein[45]. The novel PTB1:34 is also curious in that it lacks the nuclear localization signal of PTB, making it exclusively cytoplasmic. Our studies of PTB1:34 are therefore quite relevant to its *in vivo* function.

PTB1:34 is a unique domain

This complex of RRM3 and RRM4 must be considered as a single protein entity. While there are no folding data that describe when during protein synthesis the two RRMs become inextricably one domain, their association must occur early during protein folding. That speculation is based on the inability of the two separate RRMs to associate *in vitro*, and on the durability of the PTB1:34 domain after its formation. Mixed individual RRMs do not associate, even over a wide variety of solution conditions and temperatures, and at high protein concentrations, suggesting that the role of the 24 amino acid linker is not merely to connect the two RRMs, but to assist in formation of the final domain. Published NMR structures[20, 22] of PTB1:34 reveal that 17 of the 24 amino acids in the linker are structured (consistent with our assignments). Energetically, tethering both RRMs decreases the entropy of their association and RRM interactions with linker residues stabilizes their interaction.

The reorganization of backbone dynamics upon RRM3/RRM4 interaction could contribute to the energetics of RRM association. Formation of the RRM3/RRM4 complex restricts RRM3 to a single structure on the chemical shift timescale and anchors large parts of the linker. At the same time, RRM4 gains intrinsic backbone dynamics. It is tempting to conclude that its gain of flexibility is possible because RRM4 has low energy barriers between its conformational states. We hypothesize that part of its role in PTB1:34 is to reduce the conformational heterogeneity of RRM3 and the linker, which could be accomplished by either by 'capturing' the correct structure or through an 'induced-fit' mechanism involving both RRMs. The result is that both RRMs are coupled

through slow (microsecond) motions that provide the energetic basis for forming PTB1:34.

Changes in protein motions have been implicated in entropic compensation for complex formation in a variety of systems. Redistribution of backbone and side-chain dynamics[48] has been observed for protein-protein interactions[7,8,46], but few have compared the backbone dynamics of the free proteins to that of their complex. Systems which report increased dynamics upon interaction of molecules are rare, but include the signal transduction protein Cdc42hs side chains[48] when binding to PBD46, and the N-terminal domain of DnaJ[7], where μs-ms motions are increased when the presence of a C-terminal unstructured region induces interaction between helices III and IV. Here, we show that formation of PTB1:34 results in a new profile of backbone dynamics that differs dramatically from that in either RRM alone. All parts of PTB1:34 now move on similar timescales and with similar amplitudes.

PTB1:34 and RNA binding

A redistribution of backbone dynamics has also been observed for protein-ligand interactions[5,49], which is likely to occur when PTB1:34 binds to unstructured polypyrimidine tracts. Certainly the flexible RNA will lose conformational flexibility, but parts of the RNA could remain free to move[31,50] and even sample bound conformations. The ubiquitin-SH3 interaction provides an example of the energetics of protein:ligand association that lead to two distinct bound conformations of both proteins[47]. By analogy, the energetics of PTB1:34-RNA interactions could result in several conformations of both

RNA and protein, and suggests a possible mechanism for PTB1:34 binding to different RNA targets. The network of interactions between two RRMs facilitates communication between the domains and alters the backbone dynamics of each RRM constituent. When RNA is bound, the same network could facilitate redistribution of backbone dynamics to compensate for unfavorable binding entropy. We propose that the new range of motions of PTB1:34 is an intrinsic component of its mechanism of ligand selection.

The extent of motions throughout PTB1:34 suggests that the entire protein could move in concert. The analysis of protein motions using anisotropic normal modes (ANM)[51] gives another picture of the global dynamics of PTB1:34, and provides a sense of how the motions might be transferred through the body of the protein. Applying ANM calculations (*www.ccbb.pitt.edu/**anm**/*) to the structure of PTB1:34[20] predicts possible modes, some of which are very local. Of those that encompass the entire domain, several suggest a flexing about the interface, and a resulting pincer motion of loops from both RRMs (Fig. 8). If these large scale motions correspond to the measured microsecond dynamics, perhaps they are part of the RNA binding mechanism to bring the RNA strand into position on the surface of the domain.

We specifically propose that the changes in the dynamics of each separate RRM upon formation of PTB1:34 contribute to its RNA binding mechanism by priming the protein for binding to a flexible RNA. Experiments have shown that PTB1:34 has a much greater affinity for unstructured polypyrimidine tracts than for short tracts in a loop or bulge[52]. Furthermore, both experiment[53] and simulation[54] have shown that RRM4

binds RNA in the context of PTB1:34. PTB1:34 should be considered as an extended RRM with a complex RNA binding site, with coordinated motions that steer the flexible polypyrimidine tracts onto its surface. Motional coupling of the C-terminal RRMs is a unique feature of this domain, and intrinsic to its function. The redistribution of protein motions upon formation of the RRM3/RRM4 domain may explain both the driving force for the complex formation as well as the RNA selection mechanism of PTB1:34.

Acknowledgements

We thank Dr Greg DeKoster for his NMR advice and assistance. This work is supported by NIH GM077231 to KBH. CMM was supported in part by NIH T32 GM008492 and by a Sigma-Cori predoctoral fellowship.

Methods

Protein production and purification: The human PTB1 gene cloned into the pET 28A vector was a generous gift from Professor D. Black (University of California, Los Angeles). PTB1:34 begins at amino acid $[Met_{334}]G_{335}Asn_{336}$ and ends at Ile_{531}. RRM4 starts at $[Met]Gly_{442}Ser_{443}$ and ends at Ile_{531}. RRM3 begins at amino acid $[Met_{334}]G_{335}Asn_{336}$ and ends at Lys_{339}. Proteins were produced in *E. Coli* BL-21 DE3 gold cells grown at 37˚C in M9 minimal media using $^{15}NH_4Cl$ for the sole nitrogen source for ^{15}N labeled proteins, and $[^{13}C]_6$-glucose as the sole carbon source for the doubly labeled proteins. At OD_{600}=0.9, 1 mM IPTG was used to induce protein overexpression. Cells were harvested after 4 hours, pelleted by centrifugation at 6500

rpm, washed with 20 mM Tris, pH 7.5, 20 mM NaCl, and 2 mM EDTA, repelleted and stored at -80˚C overnight. Cells were resuspended in Buffer B (20mM Sodium Acetate, pH 5.3, 50-200 mM NaCl, 2 mM EDTA) with 20 μg/mL PMSF, 50 units/g DNase II, and Sigma protease inhibitor cocktail, and lysed using a French press. The lysate was spun down at 15,000 rpm and dialyzed against 1 L buffer B at 4˚C for 3 hours. The dialysis product was cleared via centrifugation and loaded onto a CM-sepharose column equilibrated in buffer C (50 mM Tris, pH 7.5 (room temperature), 10-100 mM NaCl). After flow-through was discarded, the proteins were eluted with a NaCl gradient in buffer C. Fractions from the column were combined and concentrated using Vivaspin centrifugal concentrators, and then exchanged into NMR buffer (20 mM potassium phosphate, pH 6.8, 100 mM KCl, 0.05% NaN_3.)

EMSA: Binding was measured using folded [α-^{32}P]RNA in 10 mM KCl and 10 μg/μL yeast tRNA (Boehringer) mixed with the purified protein constructs (10 nM – 5 μM) in 10 mM sodium cacodylate, pH 7.5, 100 mM NaCl, 1 mM $MgCl_2$, and 20 μg/mL BSA. All reactions were incubated for 30 minutes at room temperature. Glycerol loading dye was added and reactions were loaded on 8% polyacrylamide gels (37.5:1 acrylamide:bis) in 50 mM Tris-HCl/Glycine buffer. Gels were run at 7 V/cm at 4˚C for 4-5 hours.

Circular Dichroism: CD spectra were collected using a Jasco-J600 spectropolarimeter and a 0.1 cm path length cuvette. Samples were 25 μM protein in 20 mM potassium phosphate, pH 6.8, and 100 mM KCl. Guanidine HCl samples were

incubated for at least 12 hours and exact GndHCl concentrations were determined using refractive index. Unfolding curves were collected in duplicate, and fit to a 2-state model using Origin software.

NMR spectroscopy: NMR spectra were acquired on Varian Unity 500, 600 and 700 spectrometers equipped with Nalorac or Varian 5 mM triple resonance probes. Protein samples contained 100 μM-1 mM protein in 20 mM potassium phosphate, pH 6.8, 100 mM KCl, 0.005% NaN_3, and 10% D_2O. All data were collected at 25°C, calibrated against 100% methanol. Data for backbone assignments were collected from standard 3D CBCA(CO)NH, HNCACB, HNCO, and HNCACO experiments, and processed using Felix (Accelyrs). Chemical shift comparisons were made using $^{1}H/^{15}N$-HSQC experiments, and chemical shift differences were calculated as a single, weighted average. R_1, R_2, $R_{1\rho}$ and $^{1}H/^{15}N$-NOE data were collected using standard methods[21] with delay times of 17, 34*, 51, 68,85,118,152* and 186 ms for R_2, and 11, 112*, 223, 335, 503, 670, 838* and 1005 ms for R_1, with starred delays collected in duplicate for error analysis. ModelFree[55] fitting was performed for data collected at 500 MHz for RRM4, 700 MHz for RRM3, and globally fit at 500 and 600 MHz for PTB1:34 (data were also collected at 700 MHz). Relaxation interference experiments were used to collect the exchange free transverse relaxation rate, η_{xy}, with delay times of 5, 10, 15, 20, 25, 30, 40 and 50 ms using a single, semi-constant time experiment[40] at 700 MHz. R_{ex} was then calculated as the difference between R_2 and $\kappa\eta_{xy}$, where κ is a constant over all residues, under the assumption that the ^{1}H-^{15}N dipolar and the ^{15}N CSA axes are co-linear. Determination of κ for RRM4 was taken as a trimmed mean of the ratio of measured

R_2/η_{xy} (at 700MHz) for residues not undergoing exchange, and was found to be 1.35. Since PTB1:34 is larger, and does not have a clear baseline of residues which do not undergo exchange, its κ was calculated as the theoretical ratio[56,57] of R_2/η_{xy} for dipolar and CSA relaxation pathways:

$$\kappa = \frac{\delta^2 + p^2}{2\delta p}$$

Where δ is the chemical shift anisotropy of the ^{15}N nucleus,

$$\delta = \frac{\gamma_N B_0 \Delta\delta_N}{3\sqrt{2}}$$

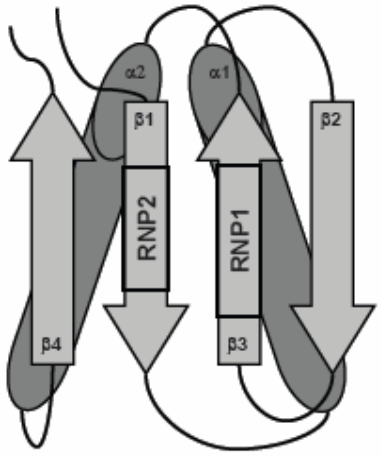
And p is the $^1H/^{15}N$ dipole-dipole coupling:

$$p = \frac{\mu_0 \gamma_H \gamma_N h}{16\pi^2 \sqrt{2} r_{HN}^3}$$

Where γ_N and γ_H are the ^1H and ^{15}N gyromagnetic ratios, B_0 is the static magnetic field strength, $\Delta\delta_N$ is the difference of the two principal components of the ^{15}N chemical shift tensor, μ_0 is the permeability of free space, h is Planck's constant, and r_{HN} is the $^1H/^{15}N$ internuclear distance.

In TRACT experiments, η_{xy} was determined from the difference in the transverse relaxation rates between the α and β spin states of all amides in the range ω_{HN} 6-10 ppm. The tumbling times were calculated assuming a rigid rotor. All relaxation data were processed using NMRPipe, and rates were calculated using NMRView (Onemoon Scientific). ModelFree[55] analysis was performed using Fast ModelFree[58], using both isotropic and axially symmetric models for calculation of rotational correlation times to determine the best fit.

a)



b)

	RNP2	RNP1
Consensus	L F V G N L I Y I K L L	K G F G F V X F R G Y A F V X Y
RRM1	I H I R K L	K N Q A F I E M
RRM2	I I V E N L	Q F Q A L L Q Y
RRM3	L L V S N L	K E N A L Q V M
RRM4	L H L S N L	R K M A L I Q M

Figure 1: Features of a canonical RRM. A typical RRM has an αβ-sandwich fold (**a**) that consists of a four stranded antiparallel β-sheet packed against two α-helices. Two RNP consensus sequences are important for protein function and reside in the center of the β-sheet, with the hexamer RNP2 sequence on β1, and the octamer RNP1 on β3. All four RRMs of PTB have RNP sequences which differ significantly from the RRM consensus (**b**). Important differences include a lack of aromatic side chains in both RNPs, which generally stack with RNA bases upon binding, as well as a lack of a glycine residue at the beginning of RNP1, thought to be important for mobility of the adjacent loop, a feature important for binding in other RRMs.

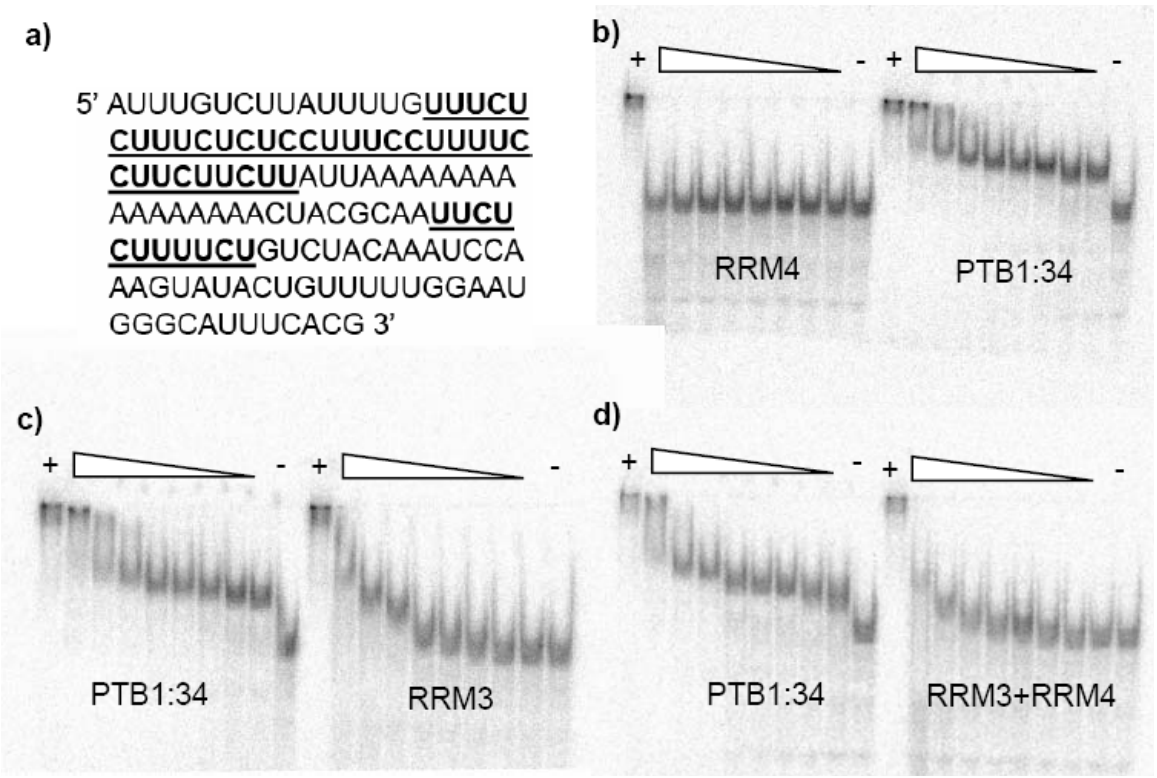


Figure 2: The binding affinity of PTB1:34 is more than the sum of its parts.

Electrophoretic mobility shift assays were used to compare the relative binding affinities of the $GABA_A$ γ2 pre-mRNA intron (**a**) to the PTB1:34 protein constructs. PTB1:34 binds at the lowest protein concentration tested, 10 nM, while RRM4 does not bind at all, even at the highest concentration tested, 5 μM (**b**). A similar comparison in (**c**) shows that RRM3 does bind to this RNA, but with around 50-fold lower affinity than PTB1:34, as the first significant band shift does not occur at protein concentrations less than 500 nM. Mixing RRM3 and RRM4 does not rescue the RNA binding (**d**), since an equimolar mixture of the two domains binds with affinity similar to that of RRM3 alone. All EMSAs were run at 4° C, and included a lane with RNA only as a negative control, and a lane with 800 nM full-length PTB, which is known to bind to this RNA with high affinity, as a positive control.

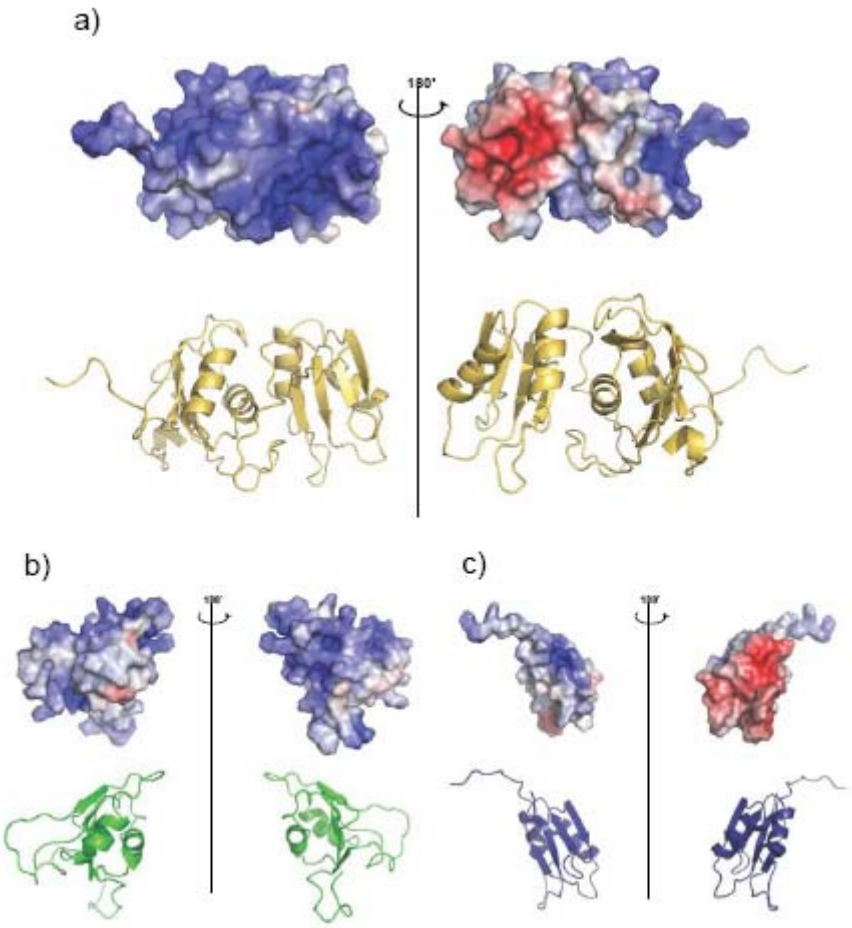


Figure 3: Surface electrostatic potentials are reorganized when PTB1:34 is formed. Electrostatic potential mapped onto the solvent accessible surface area of PTB1:34 (**a**), RRM3 (**b**), and RRM4 (**c**), with positive patches shown in blue, and negative patches in red, shows that the interaction between RRM3 and RRM4 organizes the charge distribution of the protein, and may be important to protein function.

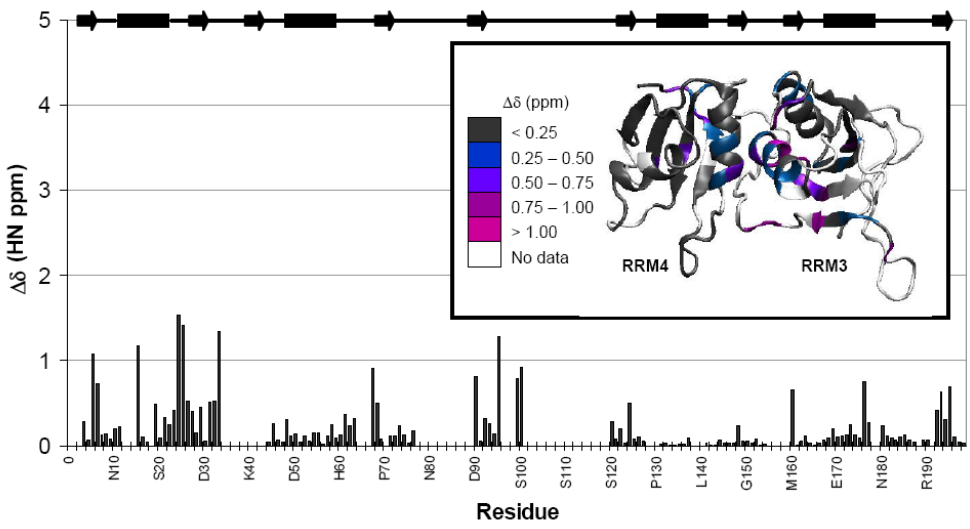


Figure 4: Changes in chemical shifts between the individual and interacting domains indicate only minor changes when the domains are separated. The change in chemical shift, Δδ, in terms of proton ppm, show that the majority of differences are concentrated at the RRM3/RRM4 interface. Δδ is shown as bars with the protein secondary structure indicated by bars (α-helix) and arrows (β-strand) along the top of the plot. For visual clarity, these changes are mapped onto the structure of PTB1:34 (inset), where white shows the areas of the protein were no data were available, grey indicates no significant Δδ (< 0.25 ppm), blue indicates Δδ between 0.25 and 0.50 ppm, violet, Δδ between 0.50 and 0.75 ppm, purple between 0.75 and 1.00 ppm, and magenta shows the most significant Δδ of greater than 1.00 ppm. Many residues in the interdomain linker are expected to have significant chemical shift changes due to altered environment but could not be calculated since the linker region was largely unassignable for the individual domains.

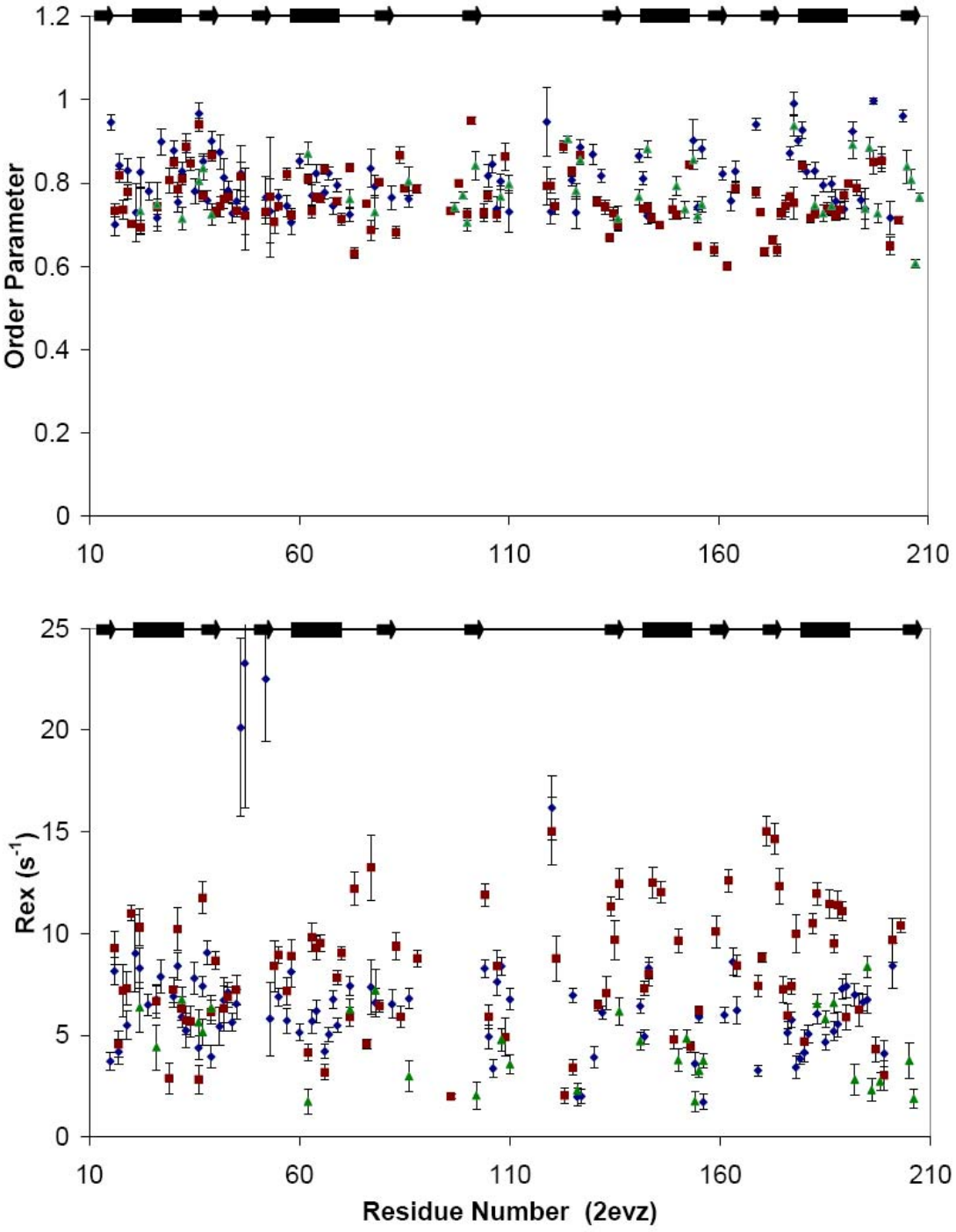

Order Parameter
1.2
1
0.8
0.6
0.4
0.2
0
10
60
110
160
210
Rex (s-1)
25
20
15
10
5
0
10
60
110
160
210
Residue Number (2evz)

Figure 5: Concentration effects do not explain the presence of pervasive Rex terms in PTB1:34. Order parameters (top panel) and R_{ex} terms (bottom panel) from ModelFree fits of 300 μM PTB1:34 at 700MHz (green triangles), 1 mM PTB1:34 at 700MHz (blue diamonds), and 1 mM PTB1:34 at 500 and 600MHz (red squares) are plotted against residue number (PDB ID: 2EVZ) with secondary structure elements indicated at the top. While there is little variation, within error, for order parameters and R_{ex} terms between the two data sets fit at 700MHz, the difference in tumbling times reflects a degree of protein self-association with τ_M= 7.2 ± 0.06 ns at 300 μM and τ_M = 9.2 ± 0.10 at 1 mM. There are much more significant differences between the fit at two static magnetic field strengths verses the fits at a single static magnetic field strength. Importantly, these data show that while under-fitting (ie: determining R_{ex} from data at a single static magnetic field strength) appears to be a problem, it is not the source of the pervasive R_{ex} terms reported. Furthermore, differences between the data sets at 300μM and 1mM at 700MHz, are localized to residues flanking the loops, suggesting that, while protein self-association may be the source of some exchange terms, it does not explain the bulk of the slow motions observed in PTB1:34.

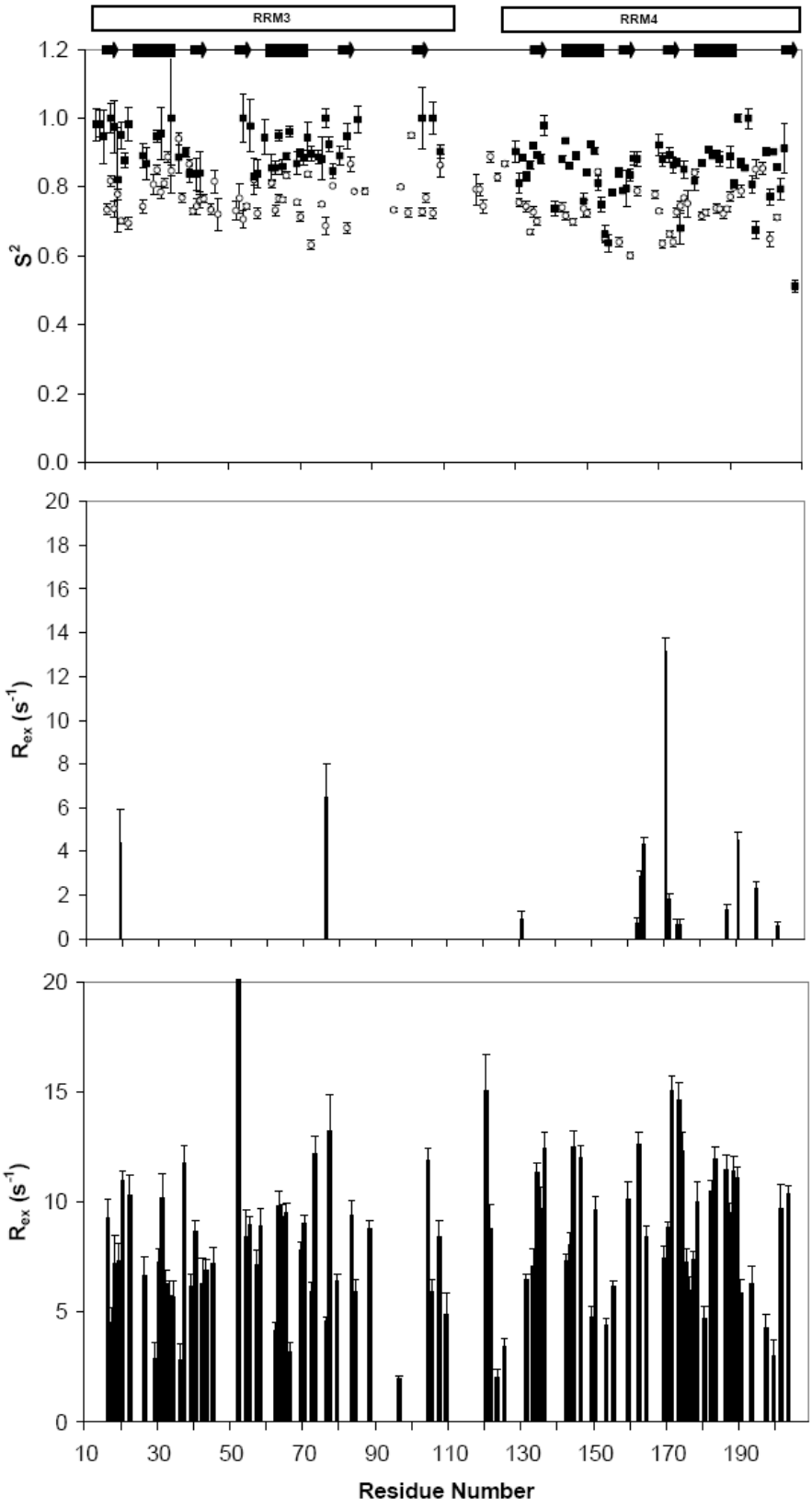

RRM3
RRM4
S²
1.2
1.0
0.8
0.6
0.4
0.2
0.0
R_ex (s^-1)
20
18
16
14
12
10
8
6
4
2
0
R_ex (s^-1)
20
15
10
5
0
10
30
50
70
90
110
130
150
170
190
Residue Number

Figure 6: ModelFree analysis suggests that slow protein motions throughout PTB1:34 occur as a consequence of the RRM3/RRM4 interaction. Lipari-Szabo order parameters, S^2, are given in the top panel for PTB1:34 (○), and the individual RRMs (■). Both RRM3 and RRM4 are much more rigid alone than in the context of PTB1:34, as evident upon comparison of the exchange contribution to transverse relaxation (R_{ex}) for PTB1:34 (bottom) and RRM3/RRM4 (middle). While PTB1:34 has uniformly dispersed R_{ex} terms of significant magnitude throughout the protein body, only a few residues in RRM3 and RRM4 require similar R_{ex} terms. This analysis shows that the differences in dynamic properties of the protein constructs are slow (μs-ms) motions that arise as a consequence of the RRM3/RRM4 interaction. Data were collected in 20 mM potassium phosphate buffer, pH 6.8, and 100 mM KCl at 500 MHz for 1 mM RRM4 and 700 MHz for 300 μM RRM3. Data for 1 mM PTB1:34 were collected at 500, 600, and 700 MHz; R_2 plots are shown in Figure S4.

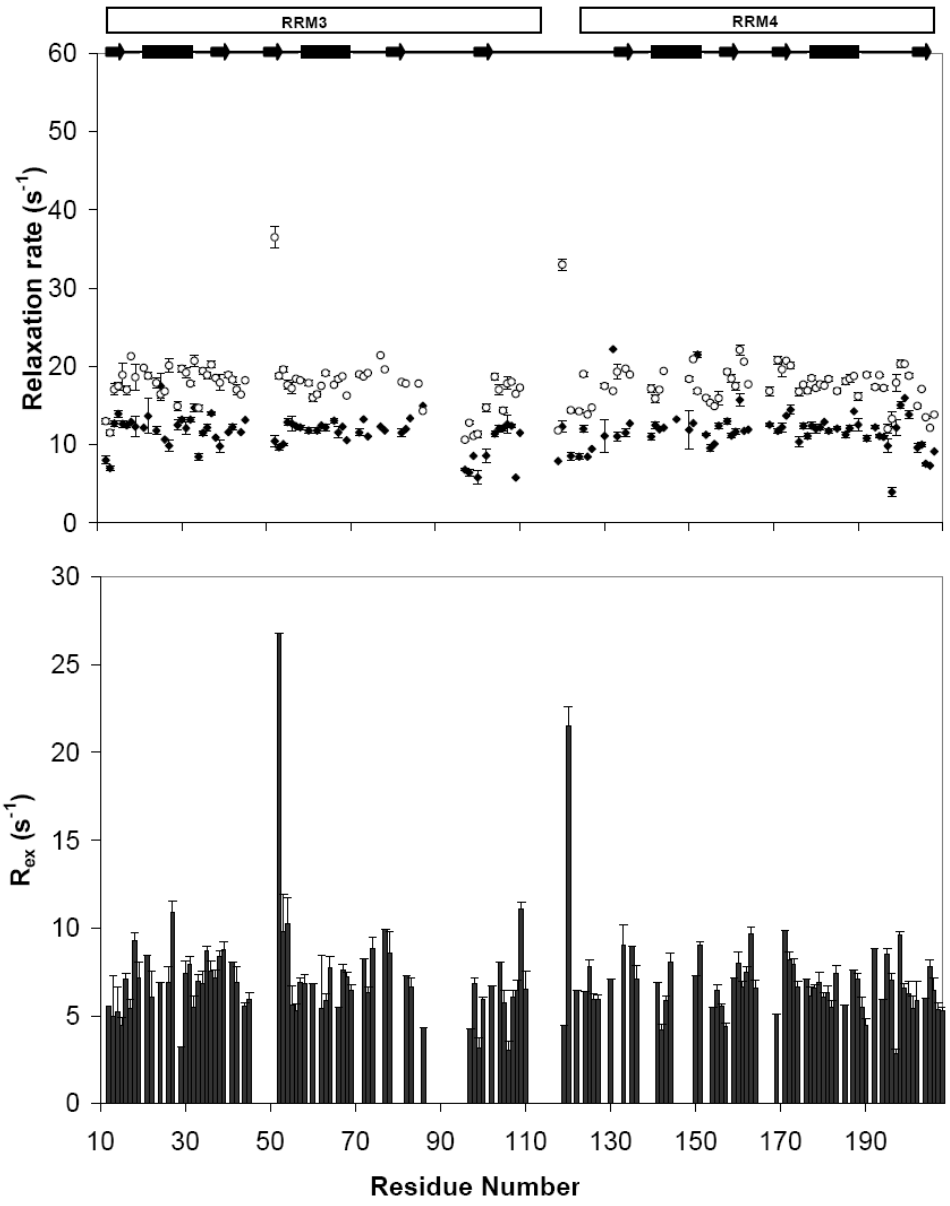
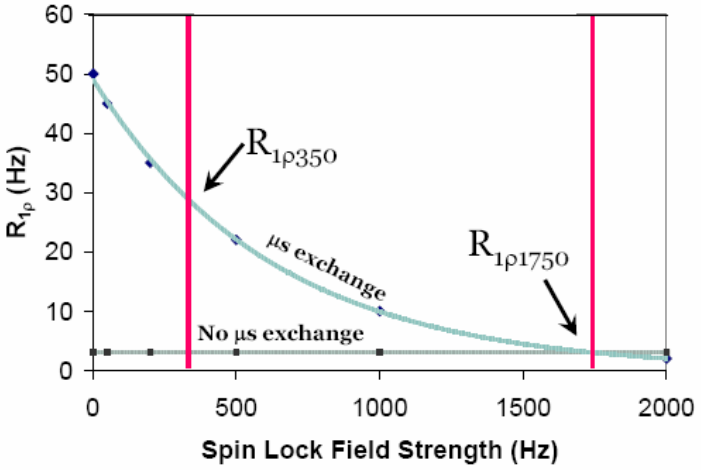
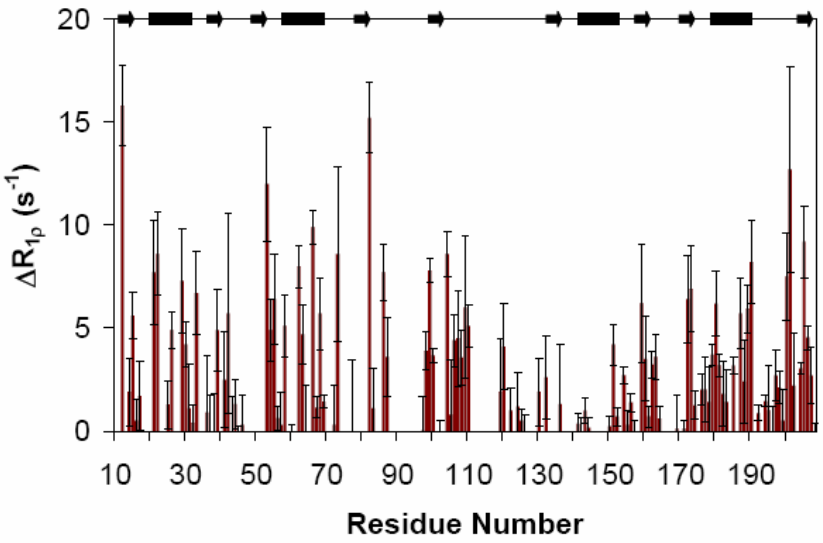

RRM3
RRM4
Relaxation rate (s-1)
0
10
20
30
40
50
60
Rex (s-1)
0
5
10
15
20
25
30
10
30
50
70
90
110
130
150
170
190
Residue Number

Figure 7: Relaxation interference experiments confirm the presence of slow motions throughout the body of PTB1:34. Transverse relaxation rates from standard experiments, R_2 (○), are compared to the exchange-free transverse relaxation rate, $\kappa\eta_{xy}$ (■), in the top panel, plotted against residue number. The difference between the two rates, R_{ex}, shown in the bottom panel, confirms the results obtained from ModelFree analysis of the relaxation data, and verify that slow motions persist throughout the body of PTB1:34. Data were collected at 700 MHz for 1 mM PTB1:34.

a)


60
50
40
30
20
10
0
$R_{1\rho}$ (Hz)
$R_{1\rho350}$
$R_{1\rho1750}$
μs exchange
No μs exchange
0
500
1000
1500
2000
Spin Lock Field Strength (Hz)

b)


20
15
10
5
0
$\Delta R_{1\rho}$ (s^{-1})
10
30
50
70
90
110
130
150
170
190
Residue Number

c)

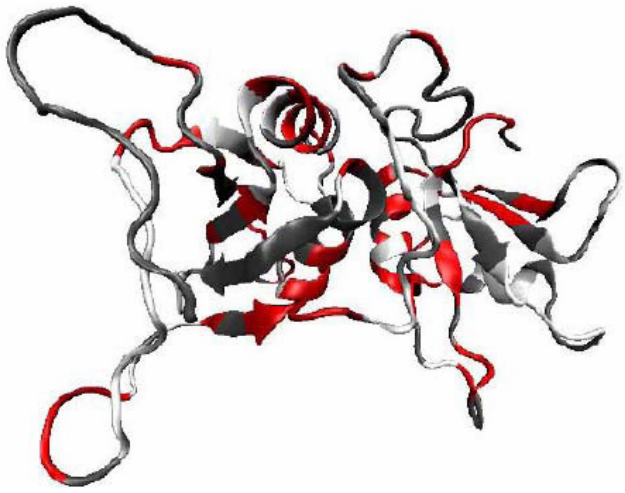


Figure 8: Residues which undergo microsecond exchange are identified by $\Delta R_{1\rho}$ experiments, and are in qualitative agreement with the ModelFree results. A schematic representation of ΔR1r is shown in (**a**) with dispersion curves shown for a residue with microsecond exchange (blue curve) and a residue with no microsecond exchange (grey curve). Importantly, this figure shows the propensity of the method to underestimate the exchange contribution to R_2 as the $R_{1\rho}$ rates for the lowest spin lock strength values are not accessible since a fit is not available to extrapolate to a spin lock field strength of 0 Hz. $\Delta R_{1\rho}$ results are shown in (**b**), plotted against residue number, and mapped onto the 3-dimensional structure in (**c**) (PDB ID: 2EVZ). These results indicate that residues throughout PTB1:34 are in microsecond exchange, and confirm that R_{ex} terms obtained from ModelFree are not fitting artifacts.

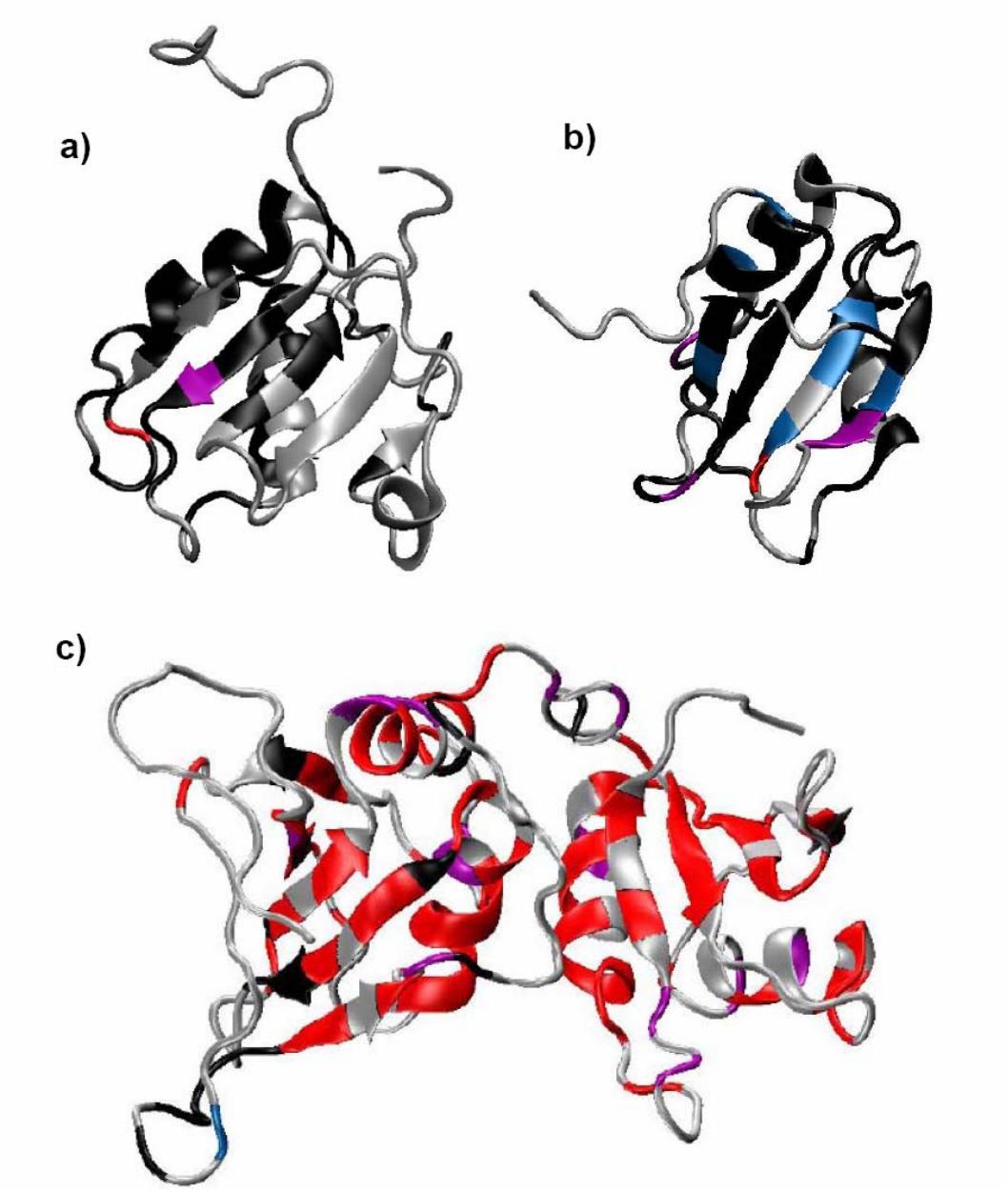

a)
b)
c)

Figure 9: The extensive slow motions throughout PTB1:34 are distinct from motions of RRM3 or RRM4 alone. Slow motions mapped onto the three dimensional structures of PTB1:34 (**a**), RRM3 (**b**) and RRM4 (**c**) (PDB ID: 2EVZ) show striking differences in dynamic properties of the constructs. Here, grey areas depict residues where no data are available, either because the residue could not be assigned or could not be fit by ModelFree; black regions show residues were data are available, but no R_{ex} term was needed to fit the data. Colored regions indicate R_{ex} terms increasing in magnitude from blue (0-2 Hz), violet (2-5 Hz) to red (> 5 Hz). While PTB1:34 has significant R_{ex} terms throughout the protein, RRM3 and RRM4 are much more rigid on this timescale, giving rise to only a handful of R_{ex} terms indicative of slow motions.

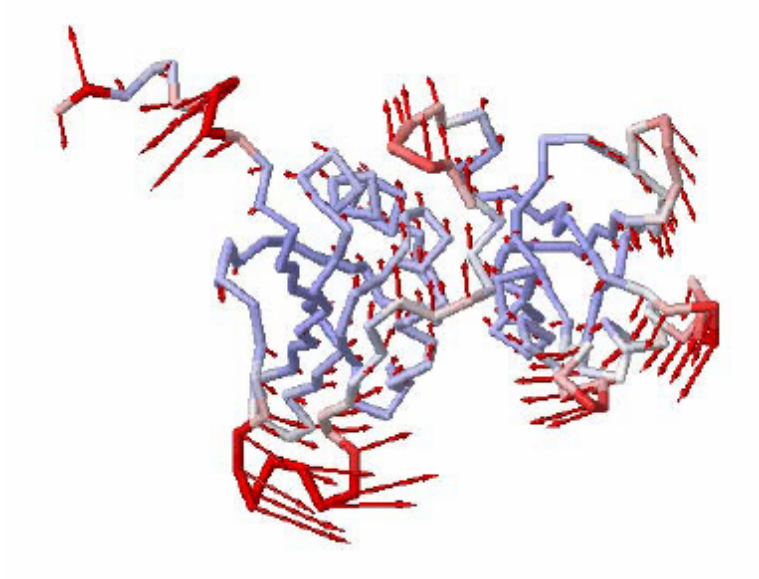


Figure 10: Normal mode calculation of the fluctuations in PTB1:34. This is mode 7 of the 20 modes calculated for structure PDB ID: 2EVZ. Red colors correspond to large fluctuations and blue colors to small fluctuations; the vectors indicate the direction of motion. RRM3 is on the left.

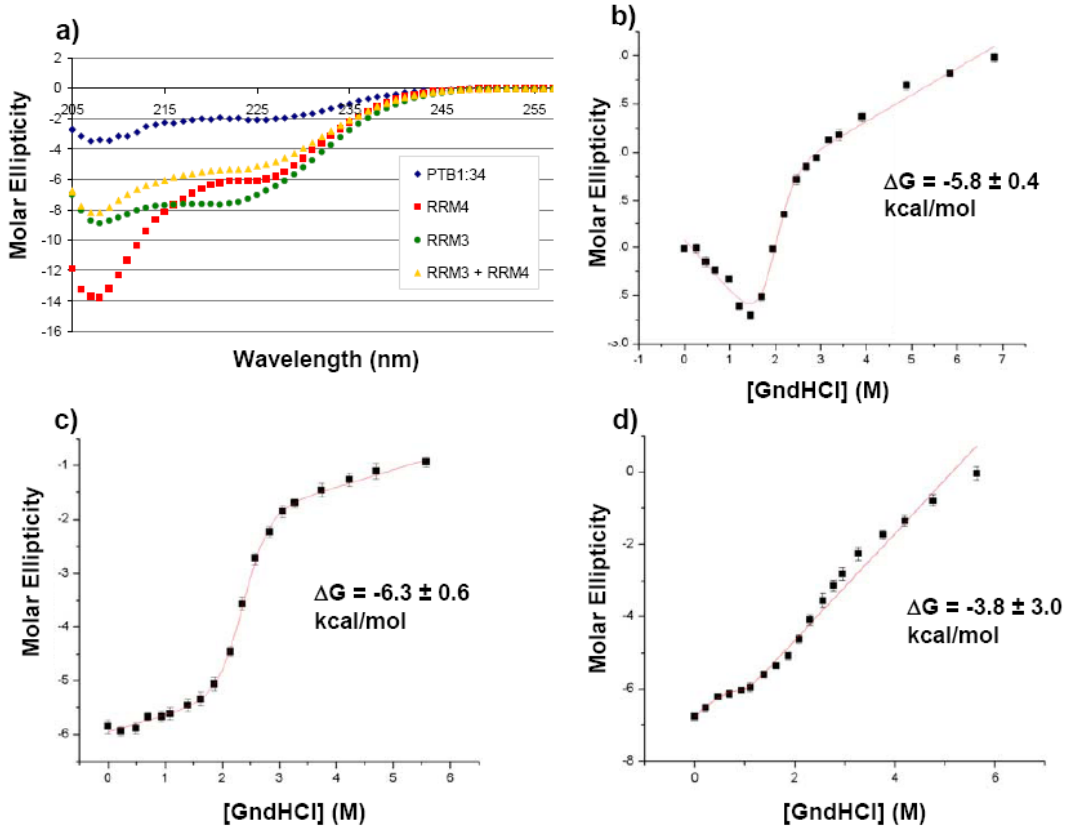


Figure S1: Secondary structure analysis. Circular dichroism (CD) spectra of each protein construct, as well as an equimolar mixture of RRM3 and RRM4 are shown in (**a**), with PTB1:34 in blue, RRM3 in green, RRM4 in red, and the RRM3/RRM4 mixture in yellow, indicating that all species are folded, and confirms that mixing the individual RRMs does not form PTB1:34. GndHCl denaturations were monitored by CD at 222 nm, and fit with a 2-state unfolding model (red lines) for PTB1:34 (**b**), RRM3 (**c**), and RRM4 (**d**). While RRM3 can be fit with this model, RRM4 and PTB1:34 cannot, indicating that the thermodynamic properties of the constructs change as a consequence of the RRM3/RRM4 interaction.

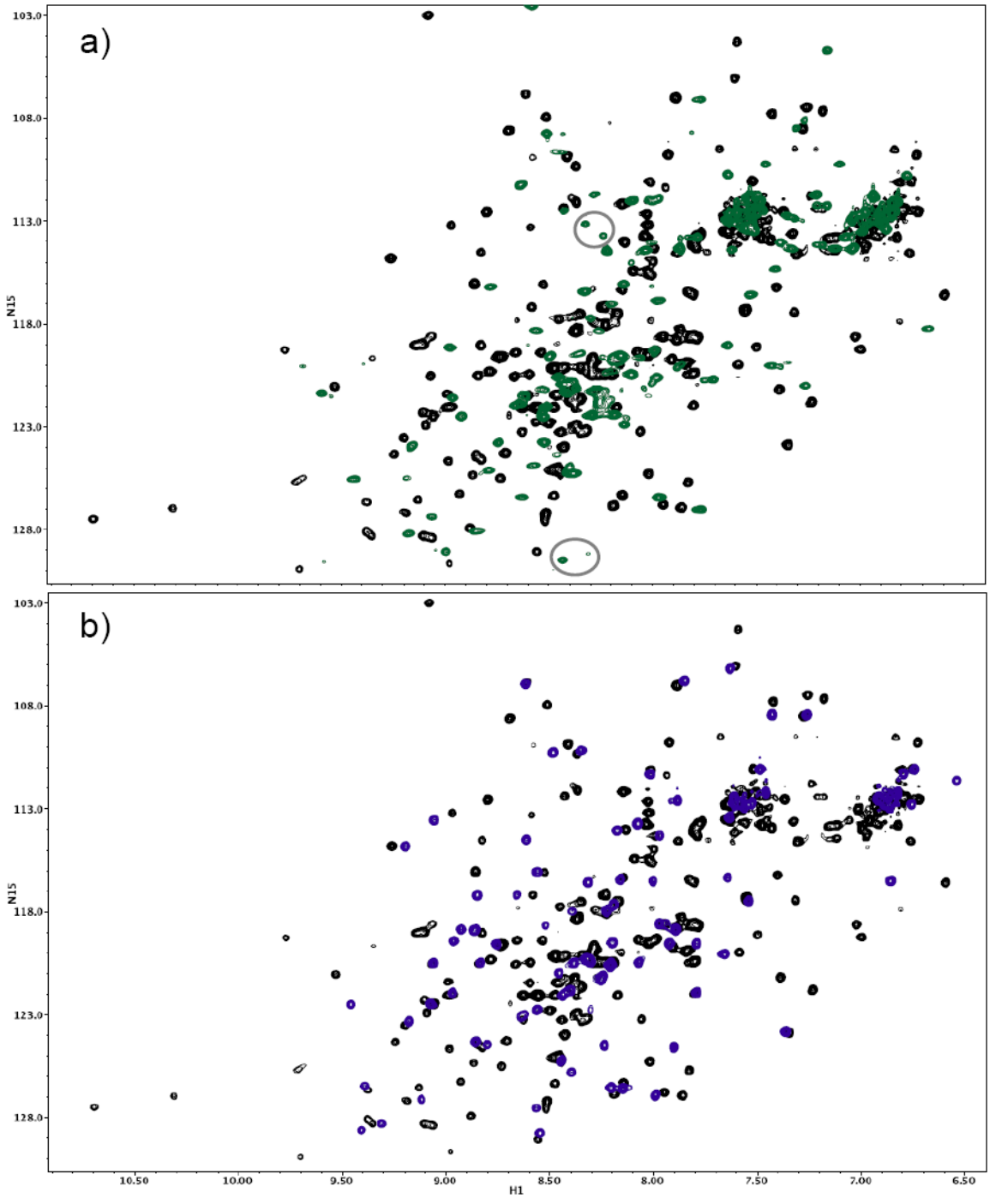

a)
b)
N15
H1
103.0
108.0
113.0
118.0
123.0
128.0
10.50
10.00
9.50
9.00
8.50
8.00
7.50
7.00
6.50

Figure S2: $^{1}H/^{15}N$-HSQC spectra comparisons show that all protein constructs are folded, but have different structural properties. A comparison of RRM3 and PTB1:34 (**a**), shows that almost none of the RRM3 peaks (green) are superimposable with the peaks from PTB1:34 (black), indicating global differences in environment. In addition, many of the RRM3 peaks are broadened, and several residues have multiple resonances (circles), indicating structural heterogeneity on the chemical shift time scale. Conversely, RRM4 (blue) has a single resonance for every amide (**b**), and the peaks are uniformly narrow. Most of the RRM4 peaks are superimposable onto the spectrum of PTB1:34. PTB1:34 displays neither the narrow line-widths of RRM4 nor the multiple resonances of RRM3, but rather behaves as a unique entity.

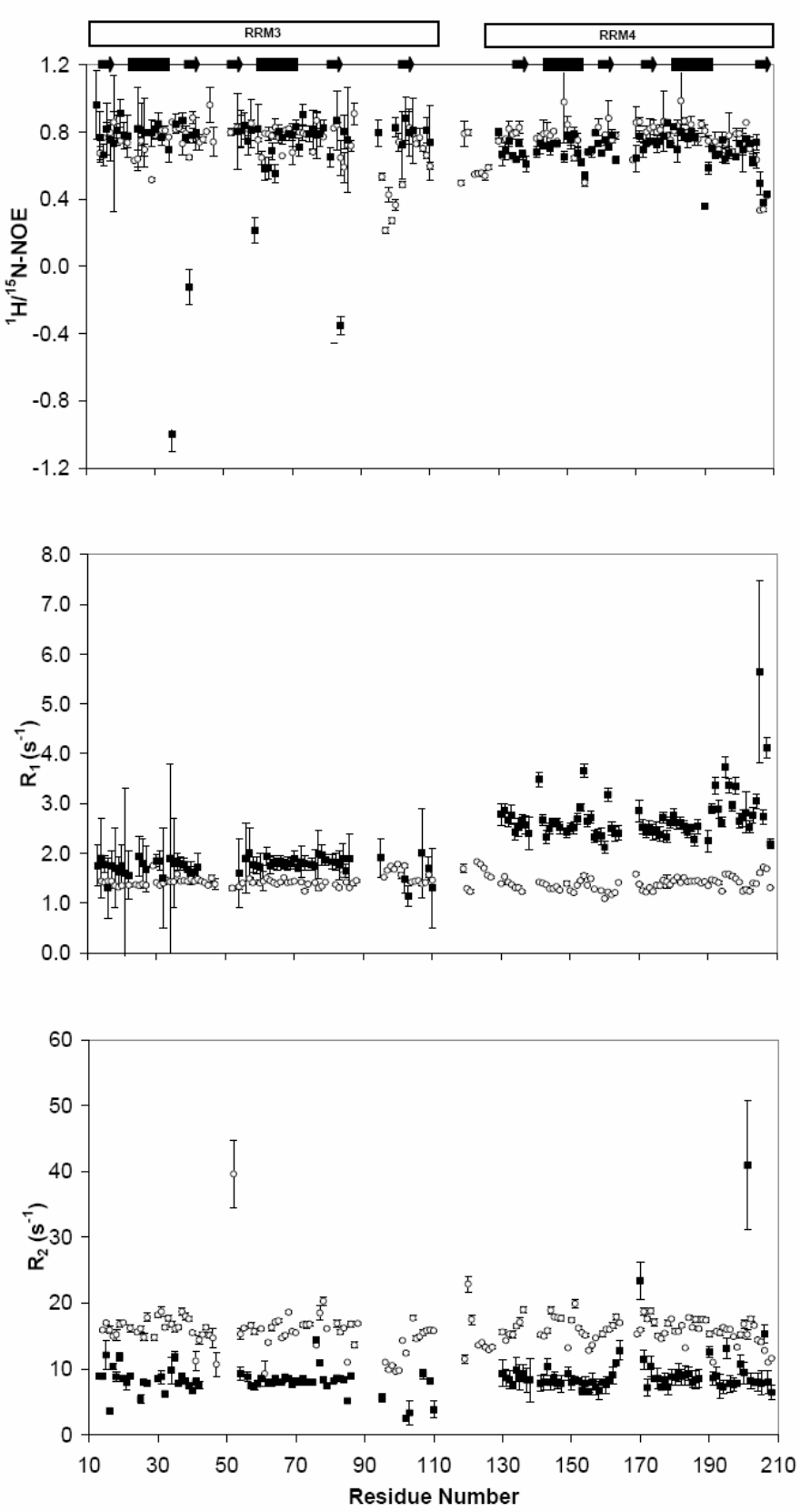

RRM3
RRM4
$^1H/^{15}N$-NOE
R_1 (s^{-1})
R_2 (s^{-1})
Residue Number

Figure S3: Differences in dynamic properties of PTB1:34, RRM3 and RRM4 are evident in the $^{1}H/^{15}N$ relaxation data. Backbone dynamics are shown for PTB1:34 (○), and individual RRMs (■), with protein secondary structure indicated along the top as bars (α-helix) and arrows (β-strand). The longitudinal relaxation rates, R_1, are displayed in the middle panel, and show a clear difference in dynamic properties between RRM3, RRM4 and PTB1:34. Importantly, PTB1:34 behaves as a single unit, even though RRM3 and RRM4 have distinctly different relaxation rates. A similar effect is observed for the transverse relaxation rates (R_2) shown in the bottom panel. Heteronuclear NOEs (top panel) indicate significant ps-ns mobility for all protein constructs, particularly in the interdomain linker, and the residues adjacent to loop3 in both RRMs. Data were collected in 20 mM potassium phosphate buffer, pH 6.8, and 100 mM KCl at 500 MHz for 1 mM RRM4, and 1 mM PTB1:34, and 700 MHz for 300 μM RRM3 and 1 mM PTB1:34.

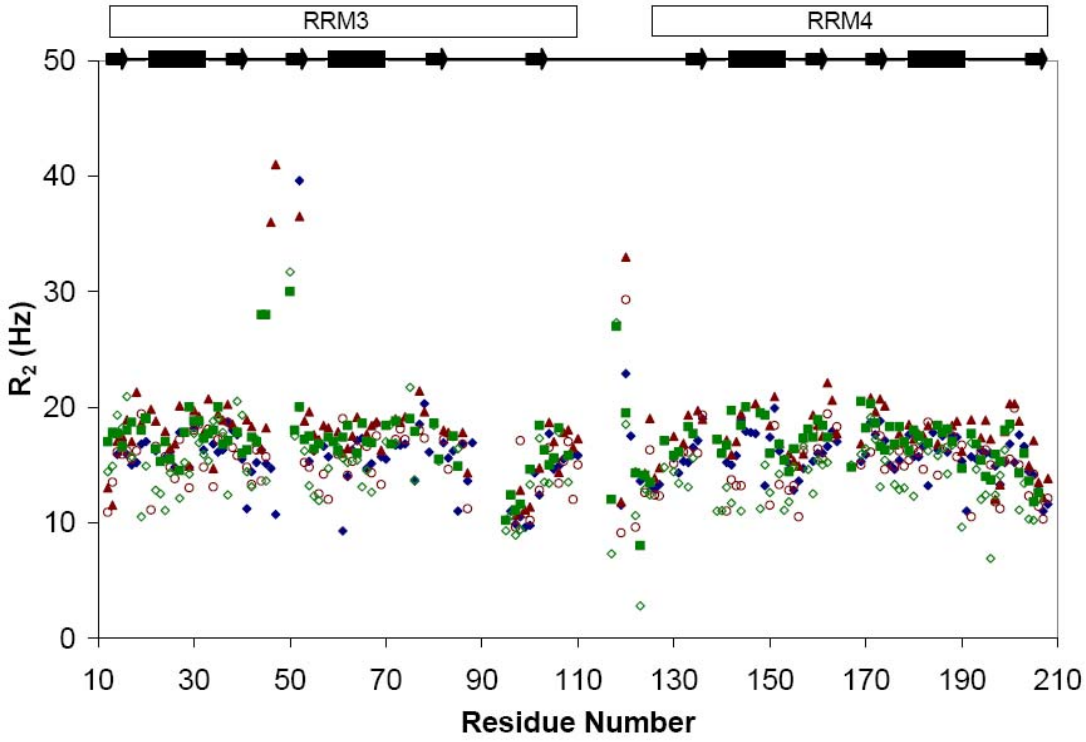


Figure S4. R_2 data for 1 mM PTB1:34 at 500, 600, and 700 MHz, 25° C. 20 mM potassium phosphate buffer, pH 6.8, and 100 mM KCl. Data collected at 500 MHz is shown in blue, 600 MHz in green and 700 MHz in red. Open symbols indicate data collected for a 300 μM protein sample.

	Concentration	τ_{MSE}	$\tau_{MModelFree}$	τ_{MTRACT}	$\tau_{MHYDRONMR}$
RRM3	0.3mM	5.2ns	5.13ns	1.5ns	7.5ns
RRM4	1.0mM	4.5ns	5.98ns	6.4ns	6.1ns
PTB1:34	0.3mM	9.7ns	7.2ns	7.0ns	32.3ns
PTB1:34	1.0mM	9.7ns	9.67ns	10.8ns	32.3ns

Table 1: Comparison of tumbling times for RRM3, RRM4 and PTB1:34. A variety of methods, including both theoretical and experimental strategies, was used to determine the tumbling times. Approximate tumbling times were calculated using the Stokes-Einstein relationship, τ_{MSE}, assuming all proteins were spherical. HYDRONMR, which calculates tumbling times based on hypothetical beads packed onto a three-dimensional structure (using coordinates from PDB ID:2EVZ), was expected to overestimate the tumbling time for flexible proteins, and gives a nonsensical result for PTB1:34. Data for 300 μM RRM3, 1 mM RRM4 as well as PTB1:34 at both 300 τM and 1 mM were collected to experimentally determine $\tau_{MModelFree}$ and τ_{MTRACT}. ModelFree fits R_2, R_1 and $^1H/^{15}N$-heteronuclear NOE data using Lipari-Szabo formalism to extract a tumbling time, and is able to accommodate both isotropic and axially symmetric models. TRACT analysis averages a decay signal from all amide protons and assumes a spherical, rigid protein.

PTB1:34	500 MHz					
Residue Number	R1	R1 error	R2	R2 error	NOE	NOE error
10						
11						
12						
13						
14	1.440	0.043	15.900	0.477	0.674	0.020
15	1.410	0.042	17.000	0.510	0.766	0.023
16	1.440	0.043	15.900	0.477	0.818	0.025
17	1.440	0.043	15.000	0.450	0.796	0.024
18	1.350	0.041	15.200	0.456	0.858	0.026
19	1.320	0.040	16.800	0.504	0.809	0.024
20	1.350	0.041	17.000	0.510	0.748	0.022
21						
22	1.360	0.041	16.200	0.486	0.736	0.022
23						
24	1.370	0.041	15.600	0.468	0.630	0.019
25	1.350	0.041	16.100	0.483	0.639	0.019
26	1.380	0.041	14.800	0.444	0.748	0.022
27	1.360	0.041	17.800	0.534	0.695	0.021
28						
29	1.750	0.053	14.800	0.444	0.515	0.015
30	1.410	0.042	18.200	0.546	0.780	0.023
31	1.360	0.041	18.700	0.561	0.811	0.024
32	1.400	0.042	16.300	0.489	0.810	0.024
33	1.460	0.044	17.700	0.531	0.809	0.024
34	1.430	0.043	16.800	0.504	0.769	0.023
35	1.430	0.043	16.100	0.483	0.859	0.026
36	1.580	0.047	16.300	0.489	0.815	0.024
37	1.440	0.043	18.700	0.561	0.824	0.025
38	1.450	0.044	18.000	0.540	0.730	0.022
39	1.440	0.043	17.600	0.528	0.791	0.024
40	1.480	0.044	15.500	0.465	0.649	0.019
41	1.440	0.043	11.200	0.336	0.885	0.027
42	1.480	0.044	14.400	0.432	0.778	0.023
43	1.440	0.043	15.200	0.456	0.739	0.022
44	1.410	0.042	16.300	0.489	0.756	0.023
45	1.360	0.041	15.100	0.453	0.803	0.024
46	1.500	0.045	14.700	0.441	0.961	0.029
47	1.380	0.041	10.700	0.321	0.741	0.022
48						
49						
50						
51						
52	1.300	0.039	39.600	1.188	0.798	0.024
53						
54	1.330	0.040	15.300	0.459	0.827	0.025
55	1.400	0.042	16.200	0.486	0.823	0.025
56						
57	1.490	0.045	16.600	0.498	0.818	0.025
58	1.410	0.042	15.700	0.471	0.851	0.026

59						
60	1.400	0.042	16.100	0.483	0.754	0.023
61	1.440	0.043	9.270	0.278	0.647	0.019
62	1.500	0.045	14.000	0.420	0.791	0.024
63	1.450	0.044	16.300	0.489	0.788	0.024
64	1.410	0.042	17.100	0.513	0.777	0.023
65	1.380	0.041	17.300	0.519	0.827	0.025
66	1.420	0.043	14.700	0.441	0.767	0.023
67	1.510	0.045	15.100	0.453	0.658	0.020
68	1.410	0.042	18.600	0.558	0.811	0.024
69	1.420	0.043	15.700	0.471	0.819	0.025
70	1.440	0.043	15.500	0.465	0.678	0.020
71						
72	1.420	0.043	16.700	0.501	0.799	0.024
73	1.240	0.037	16.700	0.501	0.793	0.024
74	1.380	0.041	16.800	0.504	0.785	0.024
75						
76	1.410	0.042	13.600	0.408	0.681	0.020
77	1.300	0.039	18.500	0.555	0.868	0.026
78	1.340	0.040	20.300	0.609	0.781	0.023
79	1.420	0.043	16.100	0.483	0.770	0.023
80						
81						
82	1.370	0.041	16.900	0.507	0.823	0.025
83	1.310	0.039	15.600	0.468	0.842	0.025
84	1.570	0.047	16.200	0.486	0.646	0.019
85	1.420	0.043	11.000	0.330	0.590	0.018
86	1.310	0.039	16.800	0.504	0.723	0.022
87	1.400	0.042	13.600	0.408	0.718	0.022
88	1.450	0.044	16.900	0.507	0.909	0.027
89						
90						
91						
92						
93						
94						
95						
96	1.520	0.046	11.000	0.330	0.534	0.016
97	1.660	0.050	9.880	0.296	0.214	0.006
98	1.740	0.052	10.500	0.315	0.426	0.013
99	1.680	0.050	9.590	0.288	0.273	0.008
100	1.780	0.053	9.750	0.293	0.365	0.011
101	1.560	0.047	14.300	0.429	0.740	0.022
102	1.750	0.053	12.400	0.372	0.486	0.015
103						
104	1.420	0.043	17.700	0.531	0.825	0.025
105	1.440	0.043	14.600	0.438	0.787	0.024
106	1.450	0.044	14.900	0.447	0.729	0.022
107	1.390	0.042	15.500	0.465	0.766	0.023
108	1.410	0.042	15.800	0.474	0.702	0.021
109	1.570	0.047	15.900	0.477	0.661	0.020

110	1.460	0.044	15.800	0.474	0.597	0.018
111						
112						
113						
114						
115						
116						
117						
118						
119	1.690	0.051	11.500	0.345	0.495	0.015
120	1.290	0.039	22.900	0.687	0.790	0.024
121	1.230	0.037	17.500	0.525	0.798	0.024
122						
123	1.830	0.055	13.600	0.408	0.548	0.016
124	1.780	0.053	14.000	0.420	0.553	0.017
125	1.700	0.051	13.300	0.399	0.556	0.017
126	1.560	0.047	12.900	0.387	0.539	0.016
127	1.520	0.046	13.300	0.399	0.589	0.018
128						
129						
130	1.390	0.042	15.600	0.468	0.746	0.022
131	1.530	0.046	14.300	0.429	0.655	0.020
132	1.420	0.043	15.300	0.459	0.758	0.023
133	1.370	0.041	15.200	0.456	0.819	0.025
134	1.310	0.039	16.500	0.495	0.785	0.024
135	1.310	0.039	17.100	0.513	0.795	0.024
136	1.230	0.037	19.000	0.570	0.826	0.025
137						
138						
139						
140						
141	1.420	0.043	15.200	0.456	0.762	0.023
142	1.400	0.042	15.000	0.450	0.767	0.023
143	1.370	0.041	15.800	0.474	0.788	0.024
144	1.290	0.039	18.900	0.567	0.792	0.024
145	1.290	0.039	17.900	0.537	0.779	0.023
146	1.330	0.040	17.800	0.534	0.804	0.024
147	1.260	0.038	17.700	0.531	0.744	0.022
148						
149	1.390	0.042	13.200	0.396	0.978	0.029
150	1.250	0.038	17.400	0.522	0.844	0.025
151	1.210	0.036	19.900	0.597	0.757	0.023
152	1.360	0.041	16.200	0.486	0.645	0.019
153	1.450	0.044	15.500	0.465	0.751	0.023
154	1.550	0.047	15.100	0.453	0.624	0.019
155	1.360	0.041	12.800	0.384	0.498	0.015
156	1.490	0.045	13.600	0.408	0.678	0.020
157	1.280	0.038	14.700	0.441	0.680	0.020
158						
159	1.250	0.038	15.300	0.459	0.828	0.025
160	1.090	0.033	16.000	0.480	0.735	0.022

161	1.220	0.037	16.000	0.480	0.783	0.023
162	1.170	0.035	16.600	0.498	0.881	0.026
163	1.210	0.036	17.900	0.537	0.751	0.023
164	1.410	0.042	17.000	0.510	0.780	0.023
165						
166						
167						
168						
169	1.580	0.047	15.300	0.459	0.634	0.019
170	1.380	0.041	15.900	0.477	0.856	0.026
171	1.250	0.038	18.700	0.561	0.860	0.026
172	1.220	0.037	17.600	0.528	0.771	0.023
173	1.310	0.039	18.800	0.564	0.800	0.024
174	1.230	0.037	17.100	0.513	0.820	0.025
175	1.360	0.041	15.000	0.450	0.830	0.025
176	1.350	0.041	14.600	0.438	0.800	0.024
177	1.460	0.044	15.400	0.462	0.812	0.024
178	1.430	0.043	16.900	0.507	0.863	0.026
179	1.370	0.041	17.600	0.528	0.780	0.023
180	1.450	0.044	15.700	0.471	0.751	0.023
181	1.520	0.046	15.700	0.471	0.806	0.024
182	1.410	0.042	16.600	0.498	0.837	0.025
183	1.510	0.045	13.200	0.396	0.985	0.030
184	1.430	0.043	17.800	0.534	0.865	0.026
185	1.440	0.043	16.300	0.489	0.797	0.024
186	1.440	0.043	17.500	0.525	0.844	0.025
187	1.450	0.044	16.100	0.483	0.772	0.023
188	1.400	0.042	17.500	0.525	0.757	0.023
189	1.430	0.043	17.400	0.522	0.846	0.025
190	1.350	0.041	15.300	0.459	0.806	0.024
191	1.470	0.044	11.000	0.330	0.747	0.022
192	1.450	0.044	15.700	0.471	0.713	0.021
193	1.410	0.042	15.500	0.465	0.742	0.022
194	1.240	0.037	16.300	0.489	0.696	0.021
195	1.580	0.047	16.000	0.480	0.661	0.020
196	1.580	0.047	16.000	0.480	0.732	0.022
197	1.540	0.046	14.900	0.447	0.779	0.023
198	1.490	0.045	13.300	0.399	0.683	0.020
199	1.420	0.043	15.200	0.456	0.788	0.024
200	1.240	0.037	16.800	0.504	0.790	0.024
201	1.270	0.038	15.200	0.456	0.731	0.022
202	1.240	0.037	17.600	0.528	0.856	0.026
203	1.400	0.042	16.600	0.498	0.700	0.021
204	1.390	0.042	14.400	0.432	0.638	0.019
205	1.610	0.048	14.100	0.423	0.634	0.019
206	1.720	0.052	12.800	0.384	0.334	0.010
207	1.690	0.051	11.000	0.330	0.340	0.010
208	1.310	0.039	11.600	0.348	0.420	0.013

PTB1:34	600 MHz					
Residue Number	R1	R1 error	R2	R2 error	NOE	NOE error
10						
11						
12	1.300	0.230	13.000	6.000		
13	1.380	0.260	11.000	5.000		
14	1.010	0.060	17.000	2.600	0.761	0.090
15	0.967	0.048	17.800	2.400	0.932	0.127
16	0.970	0.070	17.700	3.300		
17	1.000	0.080	16.600	3.300	0.858	0.101
18	0.950	0.140	18.000	8.000	0.787	0.103
19	0.930	0.090	18.700	4.800	0.967	0.108
20						
21	1.010	0.210	18.000	11.000		
22	0.940	0.100	19.000	6.000	0.842	0.163
23						
24	1.000	0.060	16.400	2.700	0.894	0.169
25	0.930	0.090	15.300	3.600	0.841	0.094
26	0.980	0.110	17.000	5.000	0.853	0.093
27	1.100	0.120	15.500	4.400	0.830	0.125
28						
29	1.230	0.100	14.500	3.000		
30	0.980	0.050	17.800	2.700	0.929	0.092
31	0.950	0.090	20.000	5.000	0.844	0.091
32	1.020	0.060	18.800	3.100	0.894	0.098
33	1.020	0.090	18.800	4.800	0.893	0.106
34	1.060	0.060	17.300	2.800		
35	0.980	0.090	17.700	4.100	1.010	0.183
36	1.090	0.060	18.000	3.200	0.928	0.130
37	1.030	0.070	20.000	4.000	0.928	0.088
38	0.940	0.070	16.800	3.200	0.830	0.088
39	1.020	0.060	17.100	3.000	0.879	0.056
40						
41	0.950	0.070	17.800	3.700		
42	1.040	0.130	16.000	6.000	0.884	0.038
43	0.990	0.060	16.300	2.400	0.883	0.087
44	1.010	0.060	17.400	2.700	0.808	0.061
45	0.980	0.080	17.000	3.800	1.040	0.115
46	1.000	0.160	28.000	20.000	0.940	0.208
47			28.000	26.000	0.946	0.222
48						
49	0.910	0.220				
50						
51						
52	0.920	0.130	30.000	16.000	0.878	0.146
53	1.070	0.210	18.000	11.000	0.688	0.120
54	1.000	0.150	20.000	9.000	0.858	0.105
55	1.011	0.041	17.200	1.900		
56	0.890	0.060	17.500	2.900		
57	1.030	0.070	16.400	3.200	1.030	0.131
58	0.940	0.090	16.800	4.100	0.831	0.079

59						
60	0.967	0.041	17.400	2.100	0.969	0.201
61	1.150	0.480	17.000	27.000	0.947	0.065
62	1.090	0.050	16.200	1.900	0.998	0.156
63	1.000	0.080	17.400	3.800	0.862	0.091
64	0.990	0.060	18.400	3.000	0.903	0.100
65						
66	0.998	0.046	16.000	1.800	0.900	0.126
67	1.016	0.038	18.600	1.900	0.856	0.119
68	0.993	0.046	17.000	2.000	0.873	0.099
69	1.011	0.043	16.900	1.800	0.942	0.104
70						
71						
72	0.980	0.050	18.400	2.500	0.850	0.052
73	0.920	0.090	16.800	4.400	0.881	0.102
74	1.001	0.049	18.800	2.500	0.869	0.093
75						
76						
77	0.950	0.140	19.000	9.000	0.781	0.098
78	0.930	0.080	17.900	4.900	0.892	0.064
79						
80						
81						
82	0.970	0.080	18.600	4.500	0.981	0.121
83	0.900	0.080	15.500	3.700	0.945	0.209
84						
85						
86	0.940	0.070	17.500	3.100	0.785	0.095
87	1.060	0.080	14.900	3.000	0.726	0.024
88						
89						
90						
91						
92						
93						
94						
95						
96						
97	1.278	0.049	10.200	1.000	0.315	0.036
98	1.219	0.026	12.400	0.700		
99	1.360	0.045	11.100	0.900	0.588	0.030
100	1.158	0.046	11.600	1.100		
101						
102	1.300	0.080	14.600	2.400	0.667	0.066
103						
104	0.980	0.060	18.400	2.900		
105	1.000	0.070	16.300	3.300	0.815	0.169
106	1.030	0.060	15.000	2.100	0.762	0.061
107	0.970	0.090	15.600	4.000	0.898	0.092
108	0.970	0.050	18.200	2.600	0.846	0.100
109						

110	1.043	0.048	15.800	1.800	0.835	0.148
111						
112						
113						
114						
115						
116						
117						
118						
119	1.190	0.230	12.000	6.000	0.598	0.027
120	0.900	0.080	27.000	8.000	0.785	0.060
121						
122	0.960	0.060	19.500	3.500		
123					0.666	0.115
124	1.330	0.060	14.300	1.600	0.656	0.084
125	0.960	0.150	8.000	9.000		
126	1.180	0.047	14.100	1.400	0.620	0.063
127	1.110	0.060	13.500	1.800	0.645	0.056
128						
129						
130	1.010	0.070	17.100	3.200	0.830	0.054
131						
132	0.980	0.050	15.800	2.000	0.831	0.081
133	0.960	0.090	16.100	4.400	0.863	0.088
134						
135	0.900	0.080	18.300	4.700	0.937	0.151
136	0.860	0.060	17.700	3.100	0.886	0.078
137						
138						
139						
140						
141	1.018	0.046	17.200	2.200	0.887	0.084
142	1.004	0.041	16.000	1.600	0.882	0.141
143	0.960	0.070	16.700	3.100	0.868	0.099
144	0.910	0.060	19.700	3.800	0.846	0.009
145						
146	0.940	0.050	18.400	2.500	0.865	0.068
147	0.830	0.060	20.000	4.000	0.852	0.102
148						
149						
150	0.900	0.060	19.500	3.200	0.877	0.029
151	0.880	0.070	19.300	4.100	0.904	0.007
152	1.003	0.049	16.000	1.900	0.861	0.148
153						
154	1.136	0.042	16.800	1.700	0.706	0.094
155	0.978	0.039	15.400	1.500	0.596	0.079
156	1.079	0.041	14.400	1.500	0.820	0.080
157	0.888	0.050	16.400	2.200	0.884	0.160
158						
159	0.890	0.090	17.300	4.500		
160	0.830	0.050	18.100	2.900	0.777	0.016

161	0.918	0.038	17.400	1.700	0.854	0.074
162	0.870	0.050	18.900	3.200	0.847	0.074
163	0.920	0.050	18.400	2.700	0.809	0.042
164	1.050	0.060	17.400	3.000		
165						
166						
167						
168						
169	1.210	0.050	14.800	1.500		
170						
171	0.910	0.060	20.500	3.700	0.863	0.067
172	0.850	0.050	18.200	2.800	0.853	0.110
173	0.910	0.070	20.300	3.900	0.842	0.060
174	0.870	0.070	18.600	4.300	0.830	0.038
175	0.920	0.070	16.600	3.300	0.835	0.063
176	0.970	0.080	16.300	3.600	0.863	0.117
177	1.034	0.043	18.300	2.100	0.931	0.131
178	1.130	0.070	16.800	3.200	0.910	0.123
179	0.990	0.037	18.300	1.800		
180	1.070	0.060	17.000	2.700	0.898	0.102
181	1.054	0.046	16.200	1.900	0.839	0.095
182	0.980	0.060	18.000	2.800	0.903	0.135
183	1.020	0.050	17.500	2.400	0.934	0.152
184						
185	1.035	0.049	16.900	2.000	0.882	0.154
186						
187	1.030	0.060	18.400	2.700	0.871	0.076
188	1.000	0.070	17.800	3.200	1.020	0.160
189	1.020	0.060	18.100	2.600	0.921	0.152
190	0.980	0.070	17.000	3.100	0.858	0.064
191					0.668	0.239
192	1.026	0.044	14.700	1.600	0.672	0.113
193					0.832	0.144
194	0.884	0.040	17.700	2.000	0.777	0.093
195	1.180	0.060	16.800	2.300	0.811	0.128
196	1.120	0.050	15.400	1.900	0.780	0.124
197	1.118	0.025	14.000	0.800		
198	1.140	0.060	13.700	1.700	0.790	0.109
199	1.000	0.080	15.900	3.500	0.869	0.135
200	1.070	0.060	15.300	2.100	0.798	0.006
201	0.870	0.110	18.000	6.000	0.941	0.223
202	0.900	0.060	18.500	3.300	0.907	0.140
203						
204	1.030	0.060	14.300	2.100	0.752	0.166
205	1.190	0.090	16.000	3.400	0.692	0.057
206	1.310	0.060	13.600	1.500	0.491	0.053
207	1.342	0.036	11.800	0.800	0.454	0.024
208	1.013	0.035	12.600	1.100		

PTB1:34	700 MHz					
Residue Number	R1	R1 error	R2	R2 error	NOE	NOE error
10						
11						
12	1.890	0.140	13.000	1.100	0.361	0.180
13	2.160	0.150	11.500	0.900	0.493	0.200
14	1.505	0.033	17.060	0.370	0.716	0.075
15	1.335	0.027	17.460	0.370	0.794	0.060
16	1.180	0.046	18.900	0.700	0.878	0.083
17	1.189	0.040	17.000	0.500	0.731	0.080
18	1.110	0.070	21.300	1.500	0.601	0.126
19	1.185	0.042	18.600	0.600	0.717	0.082
20						
21	1.140	0.110	19.800	1.700	1.113	0.204
22	1.420	0.060	18.800	0.800	0.996	0.120
23						
24	1.107	0.031	17.890	0.420	0.728	0.076
25	1.110	0.037	16.440	0.460	0.662	0.088
26	1.190	0.050	16.800	0.700	0.847	0.111
27	1.325	0.045	20.100	0.800	0.935	0.109
28						
29	2.070	0.080	14.900	0.800	0.507	0.098
30	1.377	0.028	19.690	0.460	0.703	0.063
31	1.175	0.036	19.200	0.600	0.723	0.078
32	1.258	0.033	17.800	0.430	0.868	0.074
33	1.258	0.049	20.700	0.700	0.788	0.097
34	2.011	0.031	14.690	0.270	0.571	0.060
35	1.130	0.045	19.400	0.700	0.737	0.085
36	1.394	0.036	18.870	0.430	0.731	0.064
37	1.290	0.029	20.220	0.430	0.920	0.067
38	1.217	0.037	18.530	0.430	0.833	0.076
39	1.351	0.036	17.910	0.420	0.755	0.065
40						
41	1.280	0.060	18.900	1.000	0.958	0.106
42	1.270	0.070	18.300	0.900	0.928	0.110
43	1.171	0.030	17.010	0.350	0.678	0.061
44	1.122	0.029	16.390	0.350	0.695	0.064
45	1.068	0.042	18.200	0.700	0.990	0.099
46	1.210	0.100	36.000	5.000	0.806	0.175
47	1.040	0.140	41.000	10.000	1.092	0.364
48						
49						
50						
51						
52	1.080	0.070	36.500	2.700	0.790	0.144
53	1.040	0.110	18.800	1.600	0.536	0.264
54	1.010	0.080	19.600	1.400	1.192	0.198
55	1.166	0.028	17.610	0.360	0.815	0.066
56	1.048	0.026	17.180	0.380	0.902	0.060
57	1.077	0.038	18.400	0.600	0.814	0.076
58	1.140	0.047	18.200	0.700	0.904	0.102

59						
60	1.219	0.023	17.860	0.260	0.828	0.051
61	1.770	0.330	16.000	6.000	0.528	0.261
62	1.671	0.032	16.420	0.360	0.734	0.066
63	1.144	0.038	17.500	0.470	0.786	0.068
64	1.180	0.032	19.140	0.410	0.939	0.070
65						
66	1.134	0.023	17.650	0.310	0.850	0.057
67	1.167	0.022	18.380	0.290	0.852	0.050
68	1.093	0.024	18.720	0.300	0.698	0.047
69	1.138	0.021	16.260	0.250	0.862	0.047
70						
71						
72	1.047	0.025	18.990	0.360	0.787	0.054
73	0.954	0.044	18.700	0.700	0.655	0.100
74	1.029	0.022	19.160	0.330	0.973	0.060
75						
76						
77	1.260	0.070	21.400	1.200	0.663	0.116
78	1.154	0.041	19.600	0.800	1.012	0.102
79						
80						
81						
82	1.114	0.041	18.000	0.600	0.914	0.094
83	0.973	0.039	17.800	0.600	0.868	0.092
84						
85						
86	1.215	0.031	17.800	0.450	0.735	0.068
87	1.610	0.050	14.300	0.700	0.693	0.089
88						
89						
90						
91						
92						
93						
94						
95						
96						
97	1.946	0.028	10.620	0.200	0.365	0.061
98	1.924	0.018	12.800	0.800	0.488	0.041
99	1.964	0.024	11.100	0.170	0.378	0.044
100	1.983	0.032	11.350	0.230	0.400	0.063
101						
102	1.957	0.045	14.720	0.440	0.682	0.093
103						
104	1.292	0.031	18.670	0.460	0.697	0.062
105	1.181	0.037	17.000	0.500	0.870	0.100
106	1.261	0.032	14.350	0.380	0.791	0.082
107	1.199	0.047	17.800	0.700	0.764	0.087
108	1.369	0.030	18.030	0.370	0.786	0.065
109	1.930	0.090	16.500	0.800	0.964	0.144

110	1.758	0.028	17.270	0.270	0.613	0.052
111						
112						
113						
114						
115						
116						
117						
118						
119	1.850	0.220	11.800	1.600	0.363	0.168
120	1.066	0.044	33.000	1.500	0.846	0.100
121						
122	1.700	0.080	14.400	0.700	0.659	0.108
123						
124	2.056	0.032	14.260	0.270	0.637	0.053
125	1.141	0.021	19.010	0.310	0.810	0.050
126	1.576	0.026	13.840	0.220	0.578	0.051
127	1.315	0.028	14.740	0.260	0.707	0.066
128						
129						
130	1.235	0.034	17.470	0.400	0.794	0.075
131						
132	1.261	0.023	16.860	0.290	0.820	0.059
133	1.027	0.050	19.300	0.800	0.815	0.094
134						
135	0.965	0.046	19.700	0.900	0.934	0.091
136	1.106	0.030	18.950	0.450	0.827	0.066
137						
138						
139						
140						
141	1.450	0.026	17.150	0.280	0.842	0.054
142	1.220	0.024	15.880	0.260	0.742	0.054
143	1.199	0.034	17.000	0.500	0.844	0.070
144	1.157	0.033	19.400	0.500	0.838	0.067
145						
146						
147			20.300	0.600		
148						
149						
150	0.970	0.022	18.390	0.340	0.991	0.065
151	0.859	0.035	20.900	0.600	0.755	0.078
152	1.545	0.026	16.840	0.280	0.779	0.054
153						
154	1.644	0.028	15.960	0.300	0.766	0.059
155	1.326	0.021	15.330	0.200	0.667	0.048
156	1.320	0.027	14.920	0.240	0.622	0.053
157	0.989	0.026	15.920	0.340	0.824	0.062
158						
159	0.965	0.046	19.300	0.900	0.771	0.090
160	0.829	0.025	18.420	0.380	0.942	0.068

161	1.249	0.024	17.480	0.330	0.789	0.061
162	0.888	0.029	22.100	0.500	0.819	0.065
163	1.089	0.033	20.600	0.500	0.832	0.064
164	1.224	0.036	17.700	0.600	0.799	0.082
165						
166						
167						
168						
169	1.494	0.022	16.820	0.210	0.783	0.043
170						
171	0.984	0.035	20.800	0.600	0.790	0.065
172	0.984	0.028	19.600	0.440	0.839	0.065
173	0.977	0.033	20.700	0.500	0.765	0.078
174	0.969	0.047	20.100	0.900	1.061	0.109
175						
176	1.090	0.037	16.750	0.470	0.872	0.079
177	1.283	0.023	17.670	0.260	0.877	0.048
178	1.568	0.042	16.900	0.500	0.832	0.077
179	1.275	0.016	18.490	0.240	0.810	0.036
180	1.347	0.028	17.230	0.330	0.847	0.061
181	1.175	0.025	17.710	0.310	0.804	0.055
182	1.116	0.026	17.530	0.340	1.005	0.071
183	1.208	0.029	18.380	0.340	0.796	0.061
184	0.986	0.031			0.754	0.074
185	1.173	0.024	16.840	0.290	0.755	0.055
186						
187	1.155	0.028	18.120	0.340	0.847	0.063
188	1.069	0.039	18.400	0.500	0.754	0.079
189	1.144	0.029	18.770	0.400	0.809	0.062
190	1.161	0.039	16.170	0.500	0.872	0.077
191						
192	1.574	0.037	18.900	0.500	0.777	0.073
193						
194	1.148	0.024	17.340	0.300	0.716	0.054
195	1.621	0.031	18.870	0.420	0.638	0.064
196	1.634	0.029	17.230	0.330	0.813	0.062
197	1.450	0.016	12.010	0.120	0.685	0.039
198	1.595	0.037	13.280	0.290	0.669	0.064
199	1.210	0.044	17.900	0.600	0.852	0.082
200	1.592	0.036	20.300	0.900	0.720	0.071
201	1.100	0.060	20.300	1.100	0.752	0.140
202	1.118	0.035	18.800	0.500	0.805	0.074
203						
204	1.311	0.029	14.930	0.300	0.729	0.068
205	1.849	0.049	17.100	0.600	0.617	0.077
206	2.067	0.037	13.490	0.300	0.529	0.064
207	1.926	0.019	12.130	0.140	0.410	0.039
208	1.140	0.018	13.820	0.200	0.476	0.046

Table S1: R_1, R_2 and $^{1}H/^{15}N$ NOE data for PTB1:34 collected at 500, 600 and 700 MHz.

Residue numbering is from PDB entry 2EVZ.

RRM3	700 MHz					
Residue Number	R1	R1 error	R2	R2 error	NOE	NOE error
10						
11						
12						
13	1.750	0.410	8.880	0.440	0.961	0.206
14	1.900	0.800	8.850	0.350	0.768	0.151
15	1.780	0.160	12.100	2.200	0.663	0.062
16	1.300	0.600	3.620	0.260	0.817	0.155
17	1.750	0.290	10.360	0.410	0.750	0.120
18	1.700	0.800	8.800	0.700	0.732	0.405
19	1.620	0.300	11.800	0.600	0.810	0.089
20	1.750	0.420	8.570	0.370	0.911	0.086
21	1.600	1.700	7.900	1.100	0.779	0.079
22	1.560	0.480	8.910	0.440	0.777	0.127
23						
24						
25	1.930	0.400	5.400	0.700	0.820	0.246
26	1.800	0.500	8.020	0.310	0.813	0.156
27	1.680	0.460	7.820	0.420	0.795	0.206
28						
29					0.794	0.079
30	1.850	0.120	8.540	0.160	0.815	0.047
31	1.850	0.200	8.800	1.000	0.849	0.065
32	1.500	1.000	6.230	0.350	0.766	0.047
33						
34	1.900	1.900	9.800	2.000	0.693	0.077
35	1.800	0.900	11.900	0.700	-0.999	0.105
36	1.840	0.170	7.800	0.500	0.846	0.058
37	1.790	0.130	8.800	0.600	-0.601	0.045
38	1.780	0.130	8.110	0.110	0.869	0.048
39	1.670	0.070	7.490	0.270	0.770	0.033
40	1.610	0.230	6.690	0.130	-0.125	0.105
41	1.620	0.110	7.900	0.800	0.780	0.045
42	1.720	0.290	7.500	0.600	0.789	0.098
43						
44						
45						
46						
47						
48						
49						
50						
51						
52						
53						
54	1.600	0.700	9.300	1.000	0.804	0.225
55					0.806	0.098
56	1.900	0.700	8.800	0.700	0.837	0.080
57	2.000	0.500	7.400	0.500	0.747	0.082
58	1.770	0.150	7.330	0.440	0.812	0.062

59	1.750	0.170	7.720	0.170	0.213	0.073
60	1.730	0.260	8.600	0.500	0.819	0.146
61						
62	1.940	0.180	7.900	0.210	0.581	0.066
63	1.750	0.170	7.840	0.190	0.585	0.067
64	1.810	0.130	8.560	0.160	0.687	0.052
65	1.790	0.120	8.020	0.220	0.553	0.053
66	1.840	0.110	8.000	0.090	0.799	0.042
67	1.780	0.090	8.740	0.170	0.766	0.035
68	1.790	0.130	8.530	0.210	-0.767	0.054
69	1.800	0.110	7.610	0.360	0.791	0.041
70	1.890	0.160	8.080	0.100	0.783	0.067
71	1.730	0.100	7.970	0.220	0.830	0.044
72	1.830	0.310	8.500	0.440	0.710	0.051
73	1.800	0.140	8.050	0.200	0.902	0.061
74						
75	1.780	0.130	7.970	0.190	0.791	0.048
76	1.740	0.320	14.400	0.500	0.814	0.117
77	1.980	0.490	10.920	0.260	0.784	0.141
78	1.970	0.210	8.290	0.180	0.792	0.080
79	1.860	0.110	7.460	0.220	0.823	0.045
80						
81	1.830	0.140	8.130	0.300	0.649	0.058
82	1.820	0.170	8.640	0.380	-0.523	0.068
83	1.760	0.290	8.550	0.340	0.871	0.176
84	1.890	0.300	8.340	0.290	-0.350	0.054
85	1.640	0.130	5.150	0.190	0.800	0.103
86	1.900	0.500	8.970	0.350	0.751	0.312
87						
88						
89						
90						
91						
92						
93						
94						
95	1.910	0.390	5.600	0.700	0.796	0.079
96						
97						
98						
99						
100					0.827	0.056
101						
102	1.480	0.290	2.500	0.500	0.722	0.201
103	1.140	0.200	3.300	1.900	0.880	0.129
104					0.797	0.143
105					0.807	0.193
106						
107	2.000	0.900	9.200	0.800		
108						
109	1.700	0.230	8.140	0.170	0.812	0.074

110	1.300	0.800	3.800	1.300	0.736	0.220
111						
112						
113						
114						
115						
116						
117						
118						
119						
120						
121						
122						
123						
124						
125						
126						
127						
128						
RRM4	**500MHz**					
Residue Number	**R1**	**R1 error**	**R2**	**R2 error**	**NOE**	**NOE error**
130	2.780	0.221	9.330	2.010	0.793	0.029
131	2.860	0.129	8.610	1.270	0.668	0.007
132	2.670	0.142	8.370	1.390	0.686	0.048
133	2.770	0.190	7.460	0.377	0.746	0.029
134	2.430	0.173	9.860	1.840	0.656	0.005
135	2.520	0.185	8.710	1.970	0.636	0.008
136	2.660	0.125	9.340	1.240	0.735	0.023
137	2.570	0.181	8.290	1.780	0.664	0.030
138	2.400	0.334	8.310	3.320	0.609	0.043
139						
140						
141	3.480	0.140	7.730	1.190	0.679	0.032
142	2.670	0.115	8.050	1.210	0.733	0.062
143	2.330	0.127	10.300	1.340	0.720	0.016
144	2.490	0.108	7.910	1.070	0.716	0.015
145	2.610	0.088	8.680	0.795	0.702	0.041
146	2.640	0.063	7.710	0.663	0.734	0.008
147	2.520	0.121	7.760	1.290	0.728	0.014
148						
149	2.420	0.096	9.250	1.030	0.651	0.034
150	2.490	0.108	7.910	1.070	0.780	0.022
151	2.510	0.130	8.390	0.280	0.750	0.051
152	2.690	0.117	8.130	0.243	0.790	0.040
153	2.920	0.081	6.760	0.752	0.676	0.028
154	3.650	0.135	7.740	1.160	0.619	0.021
155	2.650	0.127	6.570	0.187	0.538	0.028
156	2.720	0.119	8.030	1.270	0.680	0.009
157	2.330	0.120	7.250	1.230	0.693	0.008
158	2.370	0.123	6.650	1.360	0.794	0.011
159	2.340	0.176	7.930	0.368	0.731	0.005

160	2.120	0.117	7.590	1.260	0.675	0.020
161	3.180	0.131	8.280	0.251	0.753	0.015
162	2.500	0.133	9.060	1.290	0.712	0.016
163	2.390	0.147	10.900	0.436	0.781	0.030
164	2.410	0.147	12.800	1.480	0.632	0.021
165						
166						
167						
168						
169						
170	2.850	0.214	23.400	2.860	0.643	0.077
171	2.530	0.143	11.400	1.420	0.774	0.031
172	2.430	0.124	7.150	1.220	0.693	0.044
173	2.510	0.145	10.400	1.600	0.729	0.011
174	2.430	0.162	8.550	0.356	0.742	0.002
175	2.470	0.167	8.560	1.590	0.743	0.004
176	2.380	0.142	7.180	0.284	0.727	0.049
177	2.720	0.105	8.330	1.090	0.744	0.003
178	2.330	0.120	7.250	1.230	0.775	0.003
179	2.620	0.118	8.800	1.200	0.856	0.017
180	2.770	0.120	8.720	1.170	0.729	0.010
181	2.610	0.114	9.470	1.410	0.832	0.007
182	2.620	0.118	8.800	1.200	0.699	0.077
183	2.540	0.126	9.010	1.390	0.801	0.010
184	2.420	0.096	9.250	1.030	0.770	0.013
185	2.510	0.077	7.930	0.832	0.761	0.055
186	2.270	0.121	8.320	1.400	0.801	0.000
187	2.540	0.148	8.540	1.430	0.764	0.045
188						
189						
190	2.250	0.217	12.600	0.826	0.357	0.008
191	2.890	0.100	8.610	0.999	0.584	0.035
192	3.360	0.163	8.950	1.360	0.700	0.018
193	2.890	0.186	7.410	1.850	0.661	0.002
194	2.620	0.096	7.280	0.957	0.667	0.006
195	3.740	0.205	13.100	1.550	0.750	0.050
196	3.360	0.144	7.730	1.220	0.641	0.025
197	2.950	0.101	7.570	0.994	0.677	0.034
198	3.350	0.165	7.830	0.292		
199	2.650	0.149	10.700	1.470	0.653	0.008
200	2.710	0.166	9.390	1.650	0.729	0.018
201	2.820	0.408	40.900	9.800	0.693	0.131
202	2.530	0.116	8.240	1.200	0.758	0.022
203	2.770	0.155	7.880	0.275	0.734	0.035
204	3.050	0.147	8.030	1.370	0.627	0.026
205	5.640	1.830	7.860	10.800	0.737	0.046
206	2.730	0.138	15.300	1.450	0.495	0.070
207	4.110	0.207	8.030	1.760	0.380	0.005
208	2.190	0.097	6.410	1.140	0.429	0.002

Table S2: R_1, R_2 and $^{1}H/^{15}N$ NOE data for individual domains. RRM3 data were collected at 700 MHz and RRM4 data were collected at 500 MHz. Residue numbering is from PDB entry 2EVZ.

References

1. Boehr, D. D., McElheny, D., Dyson, H. J., Wright, P. E., The dynamic energy landscape of dihydrofolate reductase catalysis. (2006). *Science* **313**, 1638-42.
2. Kempf, J. G., Jung, J., Ragain, C., Sampson, N. S., Loria, J. P. Dynamic requirements for a functional protein hinge. (2007) *J. Mol. Biol.* **368**, 131-49.
3. Eisenmesser, E. Z., et al. Intrinsic dynamics of an enzyme underlies catalysis. (2005) *Nature* **438**, 117-121.
4. Labeikovsky, W., Eisenmesser, E. Z., Bosco, D. A., Kern, D., Structure and dynamics of pin1 durin catalysis by NMR. (2007) *J. Mol. Biol.* **367**, 1370-81.
5. Bracken, C., Carr, P. A., Cavanagh, J., Palmer, A. G., Temperature dependence of intramolecular dynamics of the basic leucine zipper of GCN4: Implications for the entropy of association with DNA. (1999) *J. Mol. Biol.* **285**, 2133-46.
6. Spyracopoulos, L., Lewis, M. J., Saltibus, L., Main chain and side chain dynamics of the ubiquitin conjugating enzyme variant human Mms2 in the free and ubiquitin-bound states. (2005) *Biochemistry* **44**, 8770-81.
7. Huang, K., Ghose, R., Flanangan, J. M., Prestegard, J. H., Backbone dynamics of the N-terminal domain in *E. Coli* DnaJ determined by ^{15}N- and ^{13}CO-relaxation measurements. (1999) *Biochemistry* **38**, 10567-77.
8. Bouguet-Bonnet, S., & Buck, M. Compensatory and long-range changes in picosecond-nanosecond main-chain dynamics upon complex formation: ^{15}N relaxation analysis of the free and bound states of the ubiquitin-like domain of human plexin-B1 and the small GTPase Rac1. (2008) *J. Mol. Biol.* **377**, 1474-87.
9. Wang, X, Tanaka Hall, TM. Structural basis for recognition of AU-rich element RNA

by the HuD protein (2001) *Nature Structural Biology* 8, 141 – 145.

10. Showalter, SA and Hall, KB. (2002). A functional role for correlated motion in the N-terminal RNA-binding domain of Human U1A protein. *J. Mol. Biol*. 322, 533-542.

11. Kranz, JK, and Hall, KB (1999). RNA recognition by the human U1A protein is mediated by a network of local cooperative interactions that create the optimal RNA binding surface. *J. Mol. Biol. 285*, 215-231.

12. Mittermaier, A, Varani, L, Muhandiram, DR, Kay, LE, and Varani, G (1999) Changes in side-chain and backbone dynamics identify determinants of specificity in RNA recognition by human U1A protein. *J Mol Biol. 294*, 967-979.

13. Showalter, SA and Hall, KB. (2004) Altering the RNA-binding mode of the U1A RBD1 protein. *J. Mol. Biol.* 335, 465-480.

14. Khanam, T, Muddashetty, RS, Kahvejian, A, Sonenberg, N, Brosius, J. 2006. Poly(A)-Binding Protein Binds to A-Rich Sequences via RNA-Binding Domains 1+2 and 3+4. RNA Biology 3:37-44.

15. Jenkins JL, Shen H, Green MR, Kielkopf CL. 2008. Solution conformation and thermodynamic characteristics of RNA binding by the splicing factor U2AF65. J Biol Chem. 283(48):33641-9.

16. Gil, A., Sharp, P. A., Jamison, S. F., Garcia-Blanco, M. A. Characterization of cDNAs encoding the polypyrimidine tract binding protein. (1991) *Genes Dev.* **5**, 1224-32.

17. Ghetti, A., Piňol-Roma, S., Michael, W. M., Morandi, C., Dreyfuss, G. hnRNP I, the polypyrimidine tract binding protein: distinct nuclear localization and association

with hnRNAs. (1992) *Nucleic Acid Res.* **20**, 3671-8.

18. Birney, E, Kumar, S, and Krainer, AR (1993). Analysis of the RNA-recognition motif and RS and RGG domains: conservation in metazoan pre-mRNA splicing factors. *Nucleic Acids Res 21*, 5803-5816.
19. Conte., M. R., et al. Structure of tandem RNA recognition motifs from polypyrimidine tract binding protein reveals novel features of the RRM fold. (2000) *EMBO J.* **19**, 3132-41.
20. Vitali, F. *et al.* Structure of the two most C-terminal RNA recognition motifs of PTB using segmental isotope labeling (2006). *EMBO J.* **25**, 150-162.
21. Simpson, P.J. *et al*. Structure and RNA interaction of the N-terminal RRM domains of PTB. (2004) *Structure* **12**, 1631-1643.
22. Oberstrass F. C. *et al.* Structure of PTB bound to RNA: specific binding and implications for splicing regulation (2005). *Science* **309**, 2054-2057.
23. Xu, R. M., Jokhan, L., Cheng, X., Mayeda, A. & Krainer, A. R. Crystal structure of human Up1, the domain of hnRNP A1 that contains two RNA-recognition motifs (1997). *Structure* **5**, 559-570.
24. Shamoo, Y., Krueger, U., Rice, L. M., Williams, K. R. & Steitz, T. A. Crystal structure of the two RNA binding domain sof human hnRNP A1 and 1.75 Å resolution (1997). *Nat. Struct. Biol.* **4**, 215-222.
25. Bae, E. *et al.* Structure and interactions of the first three RNA recognition motifs of splicing factor prp24 (2007). *J. Mol. Biol*. **367**, 1447-1458.
26. Gontarek R. R., et al. hnRNP C and polypyrimidine tract binding protein specifically interact with the pyrimidine-rich region within the 3'NTR of the HCV RNA

genome. (1999) *Nucleic Acids Res.* **27**, 1457-63.

27. Ashiya, M. & Grabowski, P. J. A neuron-specific splicing switch mediated by an array of pre-mRNA repressor sites: evidence of a regulatory role for the polypyrimidine tract binding protein and a brain-specific PTB counterpart (1997). *RNA* **3**, 996-1015.

28. Clerte, C., Hall. K. B. Characterization of multimeric complexes formed by the human PTB1 protein on RNA (2006). *RNA* **12**, 457-475.

29. Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. USA* 98, 10037-10041 2001.

30. Garcia-Mira, M. M., Dadqi, M., Fischer, N., Sanchez-Ruiz, J. M., Munoz, V. Experimental identification of downhill protein folding (2002). *Science* **298**, 2191-2195.

31. Showalter, W., A., & Hall, K. B. Correlated motions in the U1A snRNA stem/loop 2/RBD1 complex. (2005) *Biophys. J.* **89**, 2046-58.

32. Lu, J and Hall, KB. (1997) Tertiary structure of RBD2 and backbone dynamics of RBD1 and RBD2 of the human U1A protein determined by NMR spectroscopy. *Biochemistry 36*:10393-10405.

33. Kay, L.E., Torchia, D. A., Bax, A. Backbone dynamics of proteins as studied by ^{15}N inverse detected heteronuclear NMR spectroscopy: Application to staphylococcal nuclease (1989). *Biochemistry* **28**, 8972-8979.

34. Garcia de la Torre, J., Huertas, M. L., Carrasco, B. HYDRONMR: Predication of NMR relaxation of globular protein from atomic-level structure and

hydrodynamic calculations. (2000) *J. Magn. Reson.* **147**, 138-146.

35. Lee, D., Hilty, C., Wider, G., Wüthrich, K. Effective rotational correlation times of proteins from NMR relaxation interference. (2006) *J. Magn. Reson.* **178** 72-76.

36. Lipari, G., Szabo, A. Mode-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity. (1982) *J Am Chem Soc* **104**, 4546-4559.

37. Lipari, G., Szabo, A. Mode-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 2. Analysis of experimental results. (1982) *J. Am. Chem. Soc.* **104**, 4559-4570.

38. Pervushin, K., Vamvaca, K., Vogeli, B., Hilvert, D. Structure and dynamics of a molten flobular enzyme. (2007) *Nat. Str. Molec. Biol*. **14**, 1202-1206.

39. Gardino, A.K., Kern, D. Functional dynamics of response regulators using NMR relaxation techniques. (2007) in *Methods Enzymology* **423**, 149 – 165.

40. Liu, Y., Prestegard, J. H. Direct measurement of dipole-dipole/CSA cross-correlated relaxation by a constant-time experiment. (2008) *J. Magn. Reson.* **193**, 23-31.

41. Sawicka, K., Bushell, M., Spriggs, K. A., Willis, A. E. Polypyrimidine tract binding protein: a multifunctional RNA-binding protein. (2008) *Biochem. Soc. Trans.* **36**, 641-7.

42. Petoukhov, M. V., et al. I. Conformation of polypyrimidine tract binding protein in solution. (2006) *Structure* **14**, 1021-7.

43. Back, S. H., et al. Translation of poliovirus mRNA is inhibited by cleavage of polypyrimidine tract binding proteins executed by poliovirus 3Cpro (2002) *J. Virol.* **76**, 2529-42.

44. Hellen, C. U. T., et al. A cytoplasmic 57-kDa protein that is required for translation of picornavirus RNA by internal ribosome entry is identical to the nuclear pyrimidine tract binding protein. (1993) *Proc. Natl. Acad. Sci.* **90**, 7642-6.

45. Hamilton, B. J., Genin, A., Cron R. Q., Rigby, W. F. C. Delineation of a novel pathway that regulated CD154 (CD40 ligand) expression. (2003) *Mol. Cell. Biol.* **23**, 510-25.

46. Arumugam, S., et al. Increased backbone mobility in β-barrel enhances entropy gain driving binding of N-TIMP-1 to MMP-3. (2003) *J. Mol. Biol.* **327**, 719-34.

47. Korzhnev, D. M., Bezsonova, I., Lee, S., Chalikian, T. V., Kay, L. E. Alternate binding modes for a ubiquitin-SH3 domain interaction studied by NMR spectroscopy. (2009) *J. Mol. Biol.* **386** 391-405.

48. Loh, A. P., Pawley, N., Nicholson, L. K., Oswald, R. E. An increase in side chain entropy facilitates effector binding: NMR characterization of the side chain methyl group dynamics in Cdc42Hs. (2001) *Biochemistry* **40**, 4590-600.

49. Ferreon, J. C., Hilser, V. J. Ligand-induced changes in dynamics in the RT loop of the C-terminal SH3 domain of Sem-5 indicate cooperative conformational coupling. (2003) *Protein Sci.* **12**, 982-96.

50. Shajani, Z, Drobny, G, Varani, G. (2007) Binding of U1A protein changes RNA dynamics as observed by ^{13}C NMR relaxation studies,. *Biochemistry* **46**:5875-5883.

51. Atilgan, AR, Durrell, SR, Jernigan, RL, Demirel, MC, Keskin, O. & Bahar, I. 2001. Anisotropy of fluctuation dynamics of proteins with an elastic network model , *Biophys. J*. **80**, 505-515.

52. Clerte & Hall. 2009. The domains of polypyrimidine tract binding protein have distinct RNA structural preferences. *Biochemistry* **48**, 2063–2074.

53. Auweter, S. D., Oberstrass, F. C., Allain, F. H.-T. Solving the structure of PTB in complex with pyrimidine tracts: An NMR study of protein-RNA complexes of weak affinities (2007). *J. Mol. Biol*. **367**, 174-186.

54. Schmid, N., Zagrovic, B., van Gunsteren W. F. Mechanism and thermodynamics of binding of the polypyrimidine tract binding protein to RNA. (2007) *Biochemistry* **46**, 6500-12.

55. Mandel, A.M., Akke, M., Palmer, A.G. Backbone dynamics of *Escherichia coli* ribonuclease H1 – correlations with structure and function in an active enzyme. (1995) *J. Mol. Biol*. **246**, 144-163.

56. Goldman, M. Interference effects in the relaxation of a pair of unlike spin-½ nuclei. (1984) J Magnet Reson 60, 437-452.

57. Tjandra, N., Szabo, A., Bax, A. Protein backbone dynamics and ^{15}N chemical shift anisotropy from quantitative measurement of relaxation interference effects. (1996) J. Am. Cehm. Soc. 118, 6986 – 6991.

58. Cole, R., Loria, J.P. FAST-Modelfree: a program for rapid automated analysis of solution NMR spin-relaxation data. (2003) *J. Biomol. NMR*. **26**, 203-213.

Chapter 6

Defining an RRM for 2010: Sequence, structure and context.

RRMs are ubiquitous proteins found in prokaryotes, viruses and eukaryotes, with over 6000 identified to date, and are estimated to make up 2% of total gene products in humans (Maris, et al., 2005). In the last twenty years, over 1350 articles pertaining to RRMs have been published, yet the definition of an RRM is almost a quarter of a century old. In light of the information that has recently become available, it is appropriate to revisit the definition of an RRM, in particular, to extend the definition to account for the importance of context for these highly versatile proteins.

Origins of the traditional RRM definition

RRMs were first identified in the 1980s as proteins that accompanied mRNA precursors (pre-mRNA) and heterogeneous nuclear RNAs (nhRNAs). Further characterization revealed a 90 amino acid consensus RNA binding domain using poly(A) binding protein (Adam et al., 1986) and hnRNP protein C (Swanson et al., 1987), associating sequence information with the original functional definition of an RRM. Within these 90 amino acids, a highly conserved octamer was found consisting of mostly aromatic and positively charged side chains, and termed the ribonucleoprotein (RNP) consensus sequence (Adam et al., 1986, Swanson et al., 1987). This octamer was referred to as RNP1 when a second, less highly conserved consensus sequence, a hexamer with a singe aromatic side-chain, was coined RNP2. Conservation of these RNP sequences is illustrated in Figure 1.

The first three-dimensional structure of an RRM was reported in 1990, when Nagai, et al. solved the crystal structure of the U1 small nuclear ribonucleoprotein A (U1A) N-terminal RRM. Since then, a barrage of RRM structures have been solved using both crystallography and NMR, with over 160 structures pertaining to RRMs deposited in the protein data bank (PDB). The common structural motif that emerges is a core structure made up of a four stranded anti-parallel β-sheet packed against two α-helices. The RNP consensus sequences are located in the center of the β-sheet (Figure 2).

Structural modifications of RRMs

Although this core structure is conserved among RRMs, extensions of many flavors have been discovered. For example, α-helix 1 of the U2AF^{35} RRM is three times longer than that of a typical RRM (Kielkopf et al., 2004). The CsF-64 and the La C-terminal RRMs have C-terminal extensions that form a third α-helix that lies across the β-sheet (Jacks et al., 2003; Perez Canadillas and Varani, 2003). Other functionally important extensions may not be obvious structurally. The Bruno protein has an N-terminal extension of ten amino acids that is disordered, yet crucial for RNA binding (Lyon et al., 2009). The U11/U12-65K C-terminal RRM has a structurally canonical core structure, but requires a 30 amino acid N-terminal extension for RNA binding, even though the aromatic side-chains in the RNP sequences are necessary for RNA binding (Netter, et al., 2009). Such extensions may contribute to RNA binding ability through direct contacts with the RNA, recruitment of other proteins that contribute to binding, or by structural stabilization of the RRMs themselves.

In addition, the RNA binding properties of many RRMs depend on the presence of other RRMs. In some cases, the presence of the auxiliary RRMs is sufficient, where in other cases, such as that illustrated in chapter five by the interaction of PTB RRM3 and RRM4, extensive interactions between RRMs may be important for biological function. Sex-lethal, Hrp1 and HuD each have two RRMs that align in parallel when bound to RNA, increasing the RNA binding surface, and enhancing RNA binding affinity (Perez Canadillas, 2006; Wang and Tanaka Hall, 2001; Handa et al., 1999). On the other hand, the first two RRMs of Prp24 and Nup35 RRMs are involved in interactions in their free forms through their β-sheet surfaces, interfering with RNA binding (Bae et al., 2007; Handa et al., 2006).

Regardless of whether the structural modification stems from extensions of a single domain or interactions with other proteins or ligands, it is clear that functionally significant alterations are common for RRMs. It is important, especially for studies with truncated or separated RRMs, to consider functional importance of portions of the protein structure that may not be included in the traditionally defined RRM core structure.

Structural extensions of PTB1:34

The unique properties of PTB1:34 enable experiments aimed to determine the importance of some specific RRM extensions. The interaction between PTB RRM3 and RRM4 is required for efficient RNA binding and increases motions throughout both domains. These results are discussed in detail in chapter five.

In addition, PTB exhibits intra-domain extensions as well. PTB RRM2 and RRM3 have C-terminal extensions that extend the RNA binding surface by packing a fifth β-strand against β2 (Simpson et al., 2004; Vitali et al., 2006). This fifth β–strand is attached to β4 by a long, flexible linker that extends across the putative RNA binding surface. We designed a truncated version of PTB RRM3 that lacks the C-terminal extension and thus resembles a canonical RRM with a four stranded β-sheet: PTB1:3Δβ5. For PTB RRM3, the extension appears to be important for RNA binding directly, and does not appear to significantly impact the three-dimensional structure, stability, or interactions with RRM4.

The PTB RRM3 extension is required for RNA binding

The truncated RRM3, PTB1:3Δβ5 was designed to contain residues 10-100 using numbering from PDB ID:2EVZ (Obertrass et al., 2005). Protein production and purification were done exactly as for RRM3, which is described in Chapter two. Electrophoretic mobility shift assays (EMSAs) were employed to assess the relative binding capabilities to the GABA intron RNA (Figure 3). This RNA was chosen since it binds more readily to RRM3 than any other RNA tested. Even at the highest protein concentration tested, 5 μM, PTB1:3Δβ5 does not bind to this RNA. Significant RNA degradation was observed at concentrations greater than 5 μM, reminiscent of contamination of the protein sample with RNases. However, protein preparations repeatedly yielded a single peak from an ion-exchange column, and were judged to be pure based on SDS-PAGE. Furthermore, addition of a commercial RNase inhibitor, SUPERase-In (Ambion) did not significantly improve results.

Comparison of the three-dimensional structure and surface electrostatic potential of RRM3 and PTB1:3Δβ5 is shown in Figure 4. We did not solve the structure of either construct used in this comparison, but instead modified the PDB file for PTB1:34, 2EVZ (Oberstrass et al., 2005) to include only the desired protein sections. NMR data comparing RRM3 alone and in the context of PTB1:34 justify this comparison for RRM3 (Maynard and Hall, 2009). However, this may not be appropriate for electrostatic potential calculations for PTB1:3Δβ5. This is because the loop connecting the fourth and fifth β-strands is held in place over the β-sheet surface by the fifth β-strand. Removal of β5 would free this loop and expose the β-sheet surface. Because of this, electrostatic potential maps are shown for three protein constructs: RRM3, PTB1:3Δβ5, and a further truncated version that is missing most of this loop that accounts for exposure of the beta sheet surface, PTB1:3Δloopβ5. These results show that rearrangement of surface electrostatics may contribute to the loss of RNA binding ability observed for PTB1:3Δβ5.

In order to ensure that the inability of PTB1:3Δβ5 to bind RNA was not due to structural instability or protein folding problems, the structure was checked using CD and two-dimensional NMR. The CD spectrum of PTB1:3Δβ5 is similar to that of RRM3 (Figure 5). A slight decrease in signal is expected since a small amount of secondary structure has been lost in the truncation. However, just the opposite is observed, with a slight increase in secondary structure for PTB1:3Δβ5. These data show that the truncated protein construct is folded, and that a loss of secondary structure (ie: inability to fold) does not explain the loss of binding ability observed for the truncated version.

To get a more detailed look at possible structural consequences of the truncation, ^{15}N-labeled proteins were prepared and ^{1}H/^{15}N-HSQC spectra were collected. A comparison of the protein backbones as an overlay of the HSQC spectra is shown in Figure 6. Some changes in chemical shift are observed where expected. Assignments of RRM3 are not complete, as noted in Chapters two and five, however, the assigned resonances that are shifted correspond to backbone amides that are likely to come in close contact with β5 or the loop between β5 and β4 in RRM3. Missing peaks correspond to the deleted β5, and also loop residues that are no longer observable due to increased degrees of freedom in the loop when it is no longer anchored by the β5. Thus, the core RRM structure of PTB1:3Δβ5 remains intact, and we can tentatively conclude that the PTB RRM3 extension is not required for protein folding or stability.

The PTB RRM3 extension most likely contributes to RNA binding via direct contacts with RNA. The structures of PTB RRMs bound to RNA have been solved (Oberstrass et al., 2006), although the short tetramer RNAs that were needed to obtain sufficient NMR peak resolution were questionable in terms of binding specificity. Our attempts to identify residues involved in RNA binding using a longer RNA were not successful. Nonetheless, we were able to observe the loss of some peaks in the ^{1}H/^{15}N-HSQC titration of PTB1:34 with this longer RNA. The disappearing peaks included residues in RRM4 and RRM3, with many residues that were affected in RRM3 located in the β5 and the β4/β5 loop. These results were consistent with the structural results from Oberstrass et al., bolstering the notion that many RNA contacts in RRM3 occur in the C-terminal extension.

Taken together, the above results indicate that the role of the PTB RRM3 extension is to enhance RNA binding through direct contacts with the RNA ligand. However, more work needs to be done to confirm this conclusion. This would require the discovery of an RNA that binds specifically to RRM3, yet with a stoichiometry that is amenable to NMR studies. Furthermore, only a single RNA has been tested. More satisfying conclusions could be drawn if a larger library of RNAs were tested for binding. Finally, other labeling schemes or fancy NMR methods would need to be used in order to obtain a full assignment set for RRM3 and PTB1:3Δβ5. This is further complicated by the fact that RRM3 tends to form non-specific aggregates at high concentrations, forcing NMR experiments to be done at low concentrations.

Despite the fact that the exact mechanism of the PTB RRM3 extension cannot be determined from these studies, it is important to appreciate the functional importance of the RRM extension. Even if the structure and sequence reveal a perfectly canonical RRM, the RNA binding properties may be drastically affected if the context of the RRM is not properly accounted for. Since this context has been shown to vary widely, a good deal of caution should be exercised when working with RRMs that have been altered or removed from their biological unit.

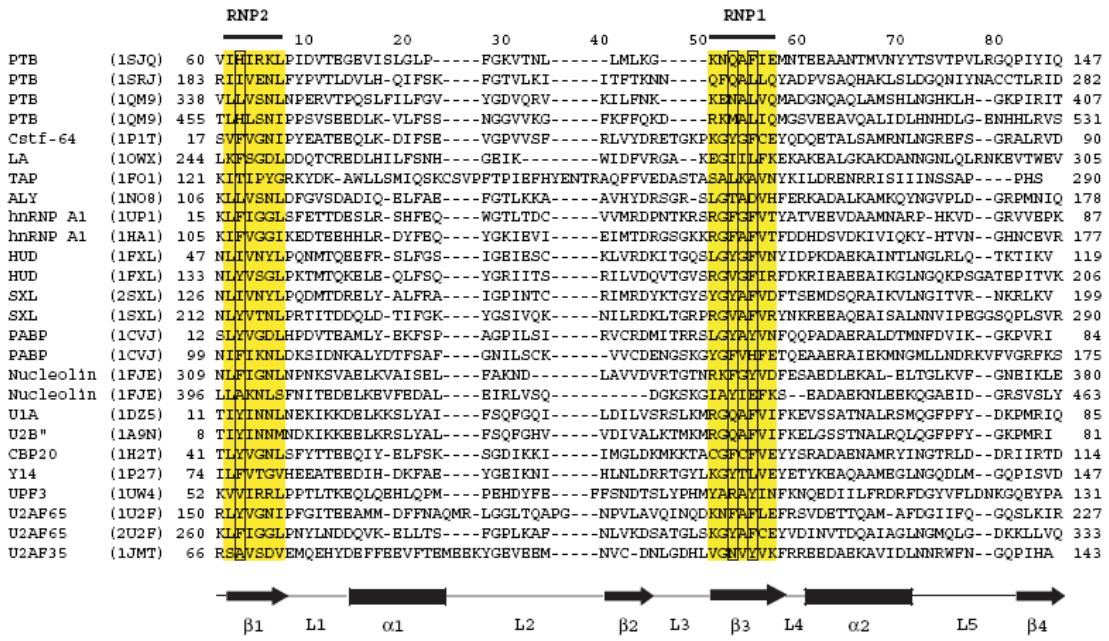


Figure 1: Conservation of RNP sequences for several RRMs RNP sequences are highlighted in yellow, with residues thought to be particularly important for RNA binding indicated. Secondary structure is shown on the bottom with β-strands as arrows, α-helices as rectangles and loops as lines. PDB ID codes are in parentheses (adapted from Maris, et al., 2005).

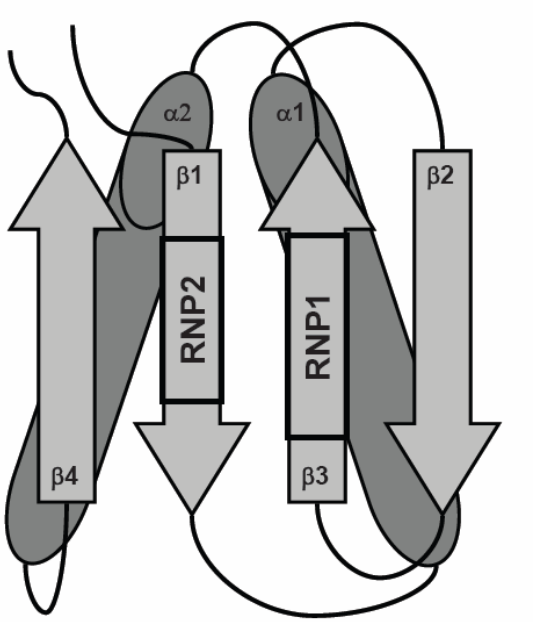


Figure 2: Cartoon of a typical RRM. A canonical RRM is made up of about 90 amino acids arranged as a four stranded anti-parallel β-sheet (arrows) packed against two α-helices (cylinders). The RNP sequences are centrally located on the β-sheet, with RNP1 on β3 and RNP2 on β1, defining the β-sheet as the RNA binding surface.

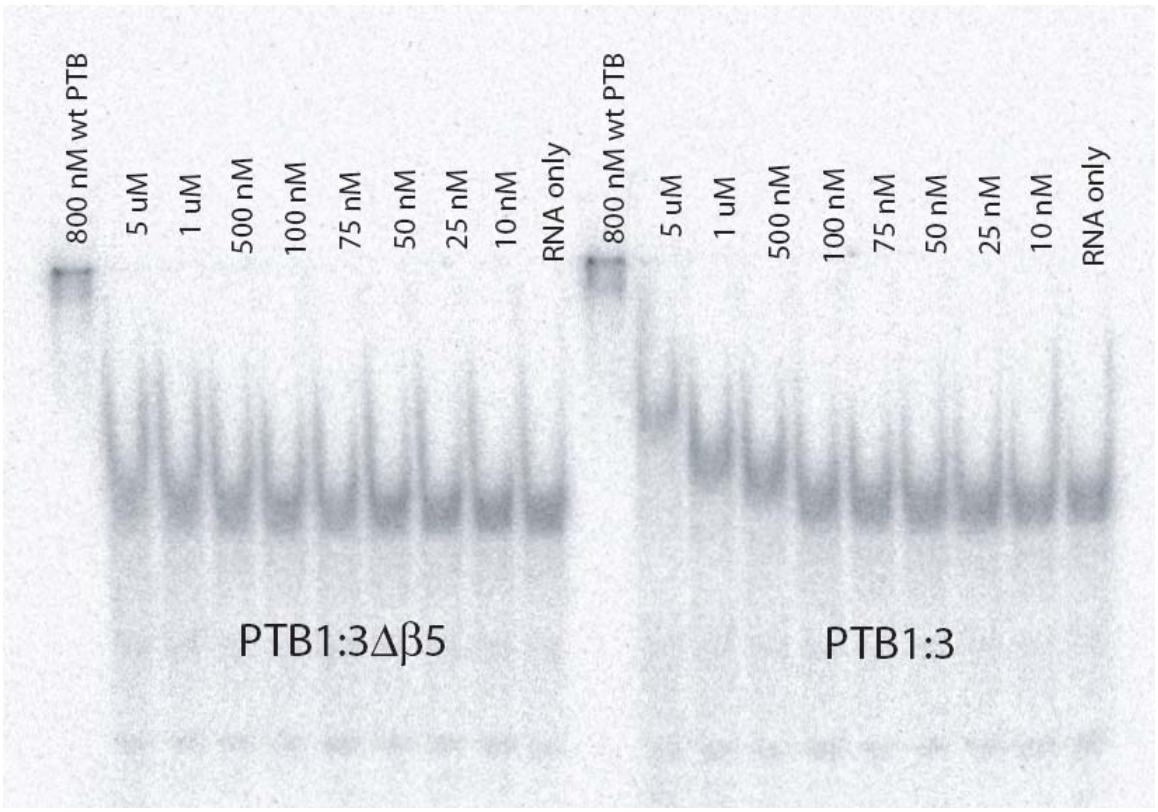


Figure 3: The PTB RRM3 extension is required for RNA binding. EMSA comparison of PTB1:34Δβ5 and PTB1:3 binding to the GABA intron RNA, which binds more readily to RRM3 than any other RNA tested, shows that RNA binding ability in the truncated version is decreased or eliminated. Binding experiments at protein concentrations greater than 5 μM are not possible due to RNA degradation.

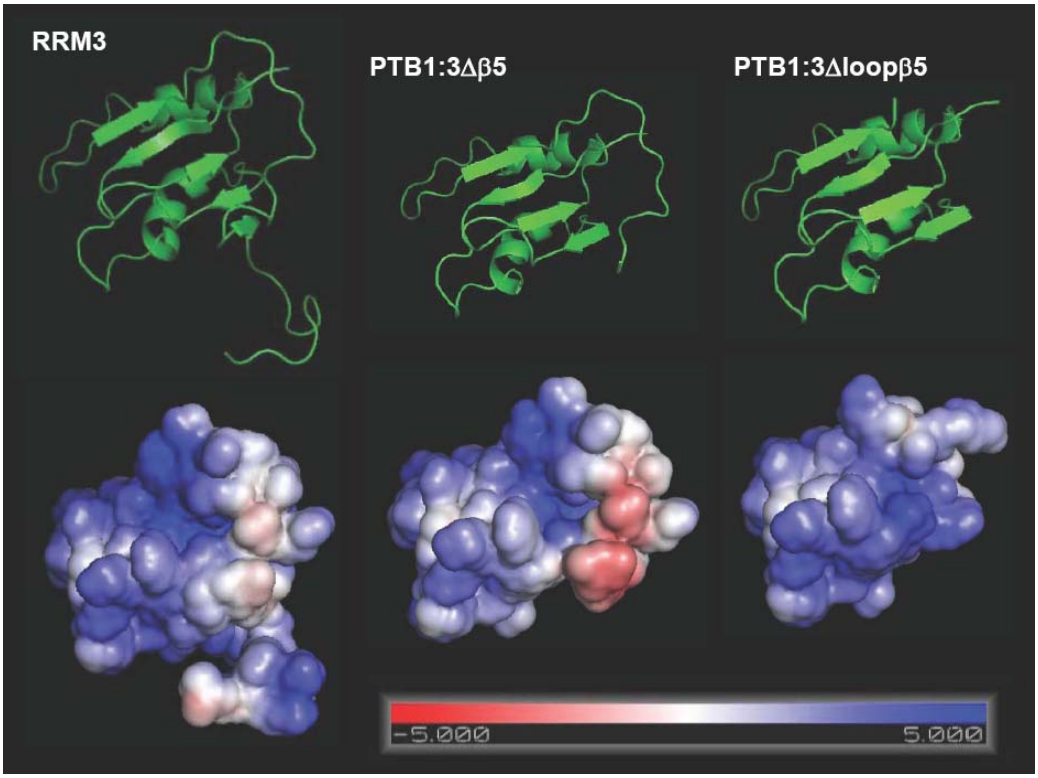


Figure 4: Truncation of the PTB RRM3 extension affects the surface electrostatic potential. The three-dimensional ribbon diagrams of RRM3, PTB1:3Δβ5 and PTB1:3Δloopβ5 are shown in the top panel, with their respective calculated electrostatic potentials mapped onto the solvent accessible surface area below. The surface of PTB1:3Δβ5 contains a small negative patch that may interfere with RNA binding. However, without β5 to anchor the loop, the negative patch probably does not remain over the β-sheet, but is more likely to move freely. The PTB1:3Δloopβ5 construct may be a better model of the surface encountered by the RNA in solution, and shows no negative patches that would interfere with electrostatic contribution to RNA binding. More detailed structural studies would be necessary to determine the actual location of the unanchored loop, and thus its influence on RNA binding.

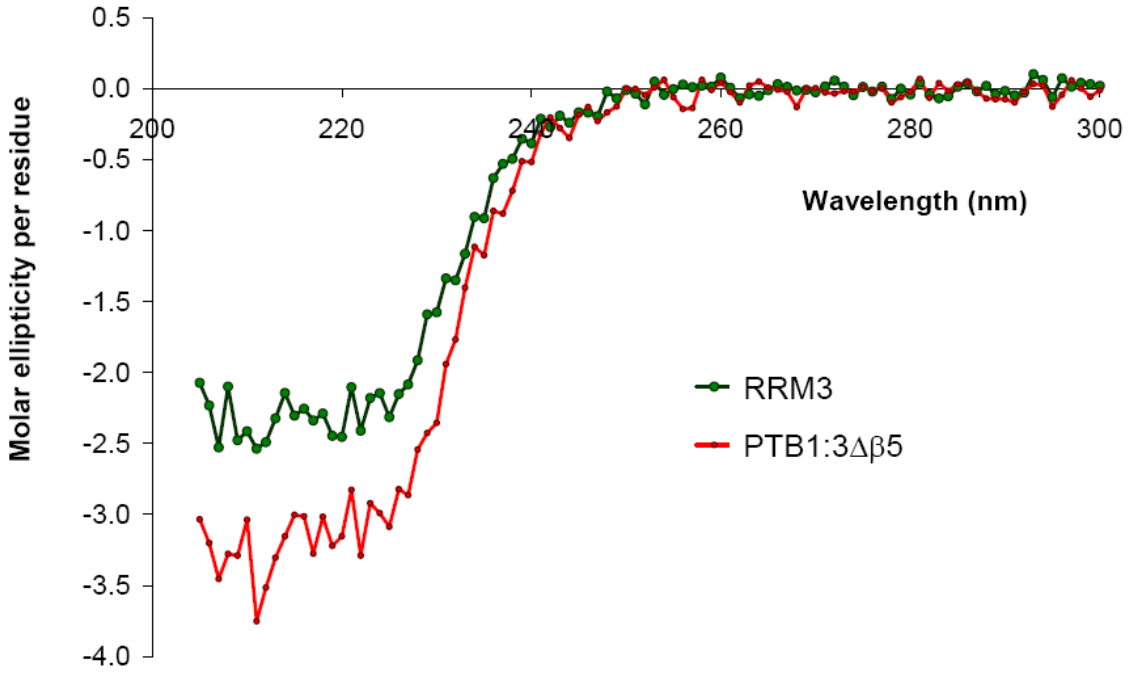


Figure 5: The PTB RRM3 extension is not necessary for protein folding. A comparison of the CD spectra for RRM3 (green) and PTB1:3Δβ5 (red) shows that the truncated RRM is folded, and even has more secondary structure than RRM3. The ellipticity has been adjusted to account for a different number of residues, so we expected the overall signal for PTB1:3Δβ5 to be equivalent to, or slightly less than that for RRM3. Nonetheless, it is clear that the truncated RRMs inability to bind RNA is not due to the protein construct being unfolded.

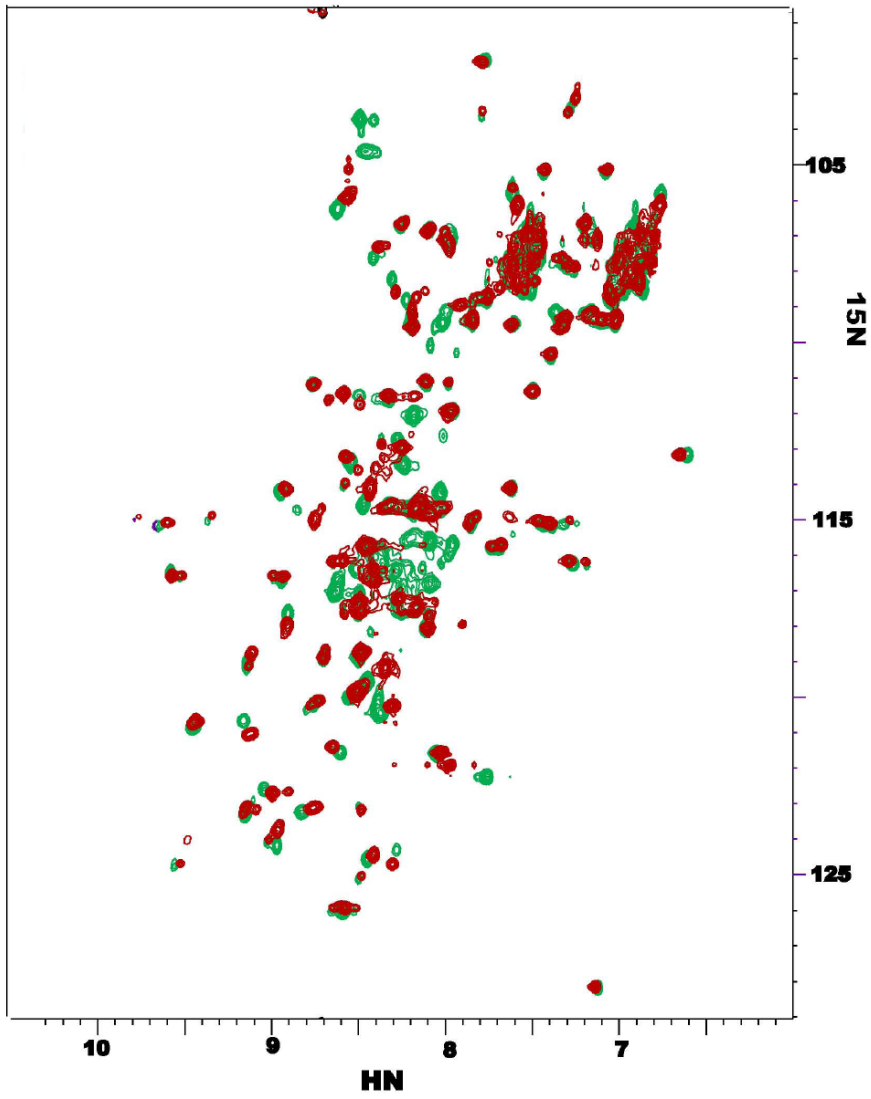


Figure 6: The overall three-dimensional structure of PTB1:3Δβ5 is very similar to that of RRM3. $^{1}H/^{15}N$-HSQC overlay of RRM3 (green) and PTB1:3Δβ5 (red) shows good chemical shift dispersion, confirming that both proteins are folded, and that there are no major structural differences between the two. Although many backbone assignments are missing for RRM3, the overall degree of differences observed, both in the number and shifting of peaks, are consistent with the changes expected from truncation of the extension in the absence of structural rearrangements.

References

Adam, S. A., Nakagawa, T., Swanson, M. S., Woodruff, T. K. and Dreyfuss, G. mRNA polyadenylate -binding -protein: gene isolation and sequencing and identification fo a ribonucleoprotein consensus sequence. (1986) *Mol. Cell Biol.*, **6**, 2932-2943.

Bae, E., Reiter, N. J., Bingman, C. A., Kwan, S. S., Lee, D., Phillips, G. N., Butcher, S. E. and Brow, D. A. Structure and interactions of the first three RNA recognition motifs of splicing factor Prp24. (2007) *J. Mol. Biol.*, **367**, 1447-1458.

Handa, N., Kukimoto-Niino, M., Akasaka, R., Kishishita, S., Murayama, K., Terada, R., Inoue, M., Kigawa, T., Kose, S., Imamoto, N., Tanaka, A., Hayashizaki, Y., Shirouzu, M. and Yokotama, S. The crystal structure of mouse Nup35 reveals atypical RNP motifs and novel homodimerization of the RRM domain. (2006) *J. Mol. Biol.*, **363**, 114-124.

Handa, N., Nureki, O., Kurimoto, K., Kim, I., Sakamoto, H., Shimura, Y., Muto, Y. and Yokoyama, S. Structural basis for recognition of the tra mRNA precursor by the Sex-lethal protein. (1999) *Nature*, **398**, 579-585.

Jacks, A., Babon, J., Kelly, G., Manolaridis, I., Cary, P. D., Curry, S. and Conte, M. R. Structure of the C-terminal domain of human La protein reveals a novel RNA recognition motif coupled to a helical nuclear retention element. (2003) *Structure*, **11**, 833-843.

Kielkopf, C. L., Lucke, S. and Green, M. R. U2AF homology motifs: protein recognition in the RRM world. (2004) *Genes Dev.*, **18**, 1513-1526.

Lyon, A. M., Reveal, B. S., Macdonald, P. M. and Hoffman, D. W. Bruno protein contains an expanded RNA recognition motif. (2009) *Biochemistry*, **48**, 12202-12212.

Maris, C., Dominguez, C. and Allain, F. H.-T. The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression. (2005) *FEBS J.*, **272**, 2118-2131.

Maynard, C. M. and Hall, K. B. Interactions between PTB RRMs induce slow motions and increase RNA binding affinity. (2009) *J. Mol. Biol.*, **397**, 260-277.

Nagai, K., Oubridge, C., Jessen, T. H., Li, J. and Evans, P. R. Crystal structure of the RNA-binding domain of the U1 small nuclear bibonucleoprotein A. (1990) *Nature*, **348**, 515-520.

Netter, C., Weber, G., Benecke, H. and Wahl, M. C. Functional stabilization of an RNA recognition motif by a noncanonical N-terminal expansion. (2009) *RNA*, **15**, 1305-1313.

Oberstrass, F. C., Auweter, S. D., Erat, M., Hargous, Y., Henning, A., Wenter, P., Reymond, L., Amir-Ahmady, B., Pitsch, S., Black, D. L. and Allain, F. H.-T. Structure of PTB bound to RNA: Specific binding and implications for splicing regulation. (2005) *Science*, **309**, 2054-2057.

Perez Canadillas, J. M. Grabbing the message: structural basis of mRNA 3'UTR recognition by Hrp1. (2006) *EMBO J.*, **25**, 3167-3178.

Perez Canadillas, J. M. and Varani, G. Recognition of GU-rich polyadenylation regulatory elements by human CstF-64 protein. (2003) *EMBO J.*, **22**, 2821-2830.

Simpson, P. J., Monie, T. P., Szendroi, A., Davydova, N., Tyzack, J. K., Conte, M. R., Read, C. M., Cary, P. D., Svergun, D. I., Konarev, P. V., Curry, S. and Matthews, S. Structure and RNA interactions of the N-terminal RRM domains of PTB. (2004) *Structure*, **12**, 1631-1643.

Swanson, M. S., Nakagawa, T. Y., LeVan, K. and Dreyfuss, G. Primary structure of human nuclear ribonucleoprotein particle C proteins: conservation of sequence and domain structures in heterogeneous nuclear RNA, mRNA, and pre-rRNA-binding proteins. (1987) *Mol. Cell Biol.*, **7**, 1731-1739.

Vitali, F., Henning A., Oberstrass, F. C., Hargous, Y., Auweter, S. D., Erat, M. and Allain, F. H.-T. Structure of the two most C-terminal RNA recognition of PTB using segmental isotope labeling. (2006) *EMBO J.*, **25**, 150-162.

Wang, X. and Tanaka Hall, T. M. Structural basis for recognition of AU-rich element RNA by the HuD protein. (2001) *Nat. Struct. Biol.*, **8**, 141-145.

Chapter 7

Unique denaturation profile of PTB RRM4 : A tool to study protein folding

The unique properties of RRM4 provide some opportunities to study protein unfolding processes. The linear denaturation profile of RRM4 (Chapter 2) allows access to states along the entire unfolding pathway. This property could be exploited to investigate several aspects of protein unfolding. Structural changes along the unfolding pathway could potentially be monitored. It would also be interesting to explore the thermodynamic changes that occur along the unfolding pathway. Finally, investigating the contribution of protein motions to the unfolding process could be quite informative.

This protein is very stable, is not sensitive to temperature, and has very well defined peaks in an NMR spectrum. In addition, we have engineered a tryptophan mutant (discussed in the paper below) in order to allow fluorescent detection of the protein. Protein production and purification schemes have already been optimized, RNA binding and biochemical characterization experiments have been completed, and NMR backbone assignments obtained. Thus, much of the preliminary work has already been done, making this system primed for some interesting experiments with a high chance of success. Too bad I spent so much time on the hard, boring stuff. Now someone else gets to do the fun part!

In addition, the linear denaturation profile is intriguing as a potential naturally

occuring downhill folder. A collabortive project addressing the downhill folding properties of RRM4 was recently published and is described in the remainder of this chapter.

A natural missing link between activated and downhill protein folding scenarios†

Feng Liu,[a] Caroline Maynard,[b] Gregory Scott,[c] Artem Melnykov,[b] Kathleen B. Hall*[b] and Martin Gruebele*[acd]

Received 27th November 2009, Accepted 11th January 2010
First published as an Advance Article on the web 11th February 2010
DOI: 10.1039/b925033f

We propose protein PTB1 : 4W as a good candidate for engineering into a downhill folder. PTB1 : 4W has a probe-dependent thermal unfolding curve and sub-millisecond T-jump relaxation kinetics on more than one time scale. Its refolding rate in denaturant is a non-linear function of denaturant concentration (curved chevron plot). Yet at high denaturant concentration its unfolding is probe-independent, and the folding kinetics can be fitted to a single exponential decay. The domain appears to fold *via* a mechanism between downhill folding and activated folding over several small barriers, and when denaturant is added, one of these barriers greatly increases and simplifies the observed folding to apparent two-state kinetics. We predict the simplest free energy function consistent with the thermal denaturation and kinetics experiments by using the singular value Smoluchowski dynamics (SVSD) model. PTB1 : 4W is a natural 'missing link' between downhill and activated folding. We suggest mutations that could move the protein into the downhill folding limit.

Introduction

The experimental detection of the protein folding 'speed limit' and of downhill folding[1–3] has attracted great interest among simulators,[4,5] theorists,[6,7] and the community interested in transition state analysis of protein folding.[8,9] A key prediction of energy landscape theory[10] is that downhill folders reach the native state without crossing a significant free energy barrier along the reaction coordinate (*i.e.* the barriers are less than 1 $RT \approx 2.5$ kJ mol^{-1}). So far, only a few small proteins have been discovered that fall into this category, as classified in ref. 11.

Strategies for finding downhill folders include: testing small proteins for probe-dependent thermodynamics;[12] designing molecules with low cooperativity but high stability;[13] and searching for wild type proteins that fold downhill.[14] Our strategy has been to extensively re-engineer already fast-folding proteins, replacing functionally conserved sidechains by those that increase stability or eliminate non-native traps. Using this approach, several small proteins have been engineered to be incipient downhill folders with barriers less than 3 RT,[15–18] downhill folders,[18,19] and even downhill folders at the melting transition.[20]

A good starting point for such engineering is a wild type protein with a broad probe-dependent melting transition (low cooperativity), yet fast folding rate (sub-millisecond). Here we present such a protein. PTB1 : 4W is a potential downhill folder with more residues and more complex topology than previously reported fast folders. It is a 91-residue protein with an α/β fold topology (Fig. 1), comprising the fourth domain of the polypyrimidine tract binding protein, an RNA binding protein involved in both pre-mRNA splicing and translation initiation.[21] The point mutation F86W was introduced in the wild type sequence so that folding thermodynamics and kinetics can be detected by tryptophan fluorescence. This mutation was conservative and led to high protein expression levels.

We characterized the equilibrium denaturation as well as the folding kinetics of this protein. PTB1 : 4W has probe-dependent thermal denaturation curves and two observed T-jump relaxation times of 18–480 μs in the absence of denaturant. Its chevron plot (log of folding rate *vs.* denaturant concentration) is curved. On the other hand, at high denaturant concentration it behaves like a two-state folder: different spectroscopic probes yield the same melting curve, and relaxation kinetics are single-exponential and slower than 10 ms. PTB1 : 4W is not quite as fast as previously reported downhill folders,[11] but it has smaller barriers than typical millisecond folders.[22] We determine the simplest free energy function compatible with the thermal melt and kinetics data, by using a new fitting approach that combines singular value-decomposed Smoluchowski dynamics (SVSD) with genetic algorithm selection of the best free energy function. PTB1 : 4W turns out to be a natural 'missing link' between activated and

[a] *Center for Biophysics and Computational Biology, University of Illinois at Urbana-Champaign, IL 61801, USA. E-mail: gruebele@scs.uiuc.edu; Fax: (217) 244 3186; Tel: (217) 333 1624*
[b] *Department of Biochemistry and Molecular Biophysics, Washington University School of Medicine, St. Louis, MO 63110, USA. E-mail: kathleenhal@gmail.com; Fax: (314) 362 7183; Tel: (314) 362 4196*
[c] *Department of Chemistry, University of Illinois at Urbana-Champaign, IL 61801, USA*
[d] *Department of Physics, University of Illinois at Urbana-Champaign, IL 61801, USA*

† Electronic supplementary information (ESI) available: Fitting parameters for the genetic algorithm. See DOI: 10.1039/b925033f

Fig. 1 Protein structure and sequence of PTB1:4. The protein structure plot was prepared with VMD.[46] Phe86 shown in purple is replaced by Trp in PTB1:4W. Comparison of the $^{1}H/^{15}N$ HSQC NMR spectra (data not shown) of the two proteins show local chemical shift changes, but no indications of significant structural changes.

downhill folding scenarios, even without sequence re-engineering. We suggest some re-engineering of its sequence that may further speed up its folding.

Results and discussion

Protein aggregation

We found that PTB1:4W has a very high propensity for aggregation when heated, like downhill-folding mutants of the lambda repressor and WW domain.[17,23] In the thermodynamic measurements, the concentration of the protein was kept at 2–3 μM to avoid aggregation up to 98 °C. Circular dichroism (CD) spectra before and after temperature titrations from 2 to 98 °C were almost identical, indicating that the measurement was reversible at low protein concentration.

Probe-dependence of protein stability

Fig. 2 shows fluorescence and CD spectra spanning the main thermal unfolding transition from 40 to 60 °C at pH 7. The fluorescence spectrum in Fig. 2B red-shifts from 342 to 352 nm upon heating, indicating solvent exposure of the tryptophan residue. The magnitude of the CD spectrum of PTB1:4W contains a significant contribution from the Trp 86 residue: native PTB1:4W has a more negative signal than PTB1:4 wild type (dotted line in Fig. 2A), even though the $^{1}H/^{15}N$ HSQC spectra of the two proteins reveal no overall structural rearrangement (data not shown). The CD signal magnitude of PTB1:4W increases upon thermal denaturation between 40 and 60 °C. This increase is due to a combination of changing aromatic side chain environment and residual structure with extended side chains. (Such 'extended structure' involves short stretches of sheet-like backbone geometry with extended side chains, but no long-range secondary or tertiary contacts.[24])

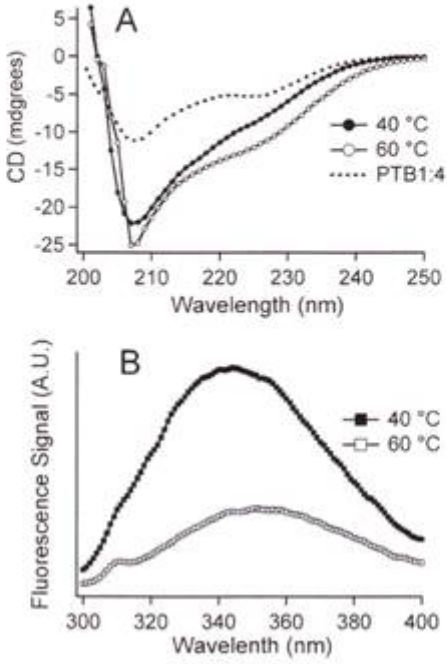


Fig. 2 (A) Circular dichroism spectra of PTB1:4W (3 μM, pH 7) spanning the main thermal denaturation transition between 40 and 60 °C. The CD spectrum of the tryptophan-free PTB1:4 (dotted line, 20 °C) is different from PTB1:4W. (The 20 °C spectrum of PTB1:4W is 20% smaller but has the same shape as the 40 °C spectrum shown.) (B) Fluorescence spectra of PTB1:4W (2 μM, pH 7) spanning the main thermal denaturation transition between 40 and 60 °C.

The normalized thermal denaturation curves monitored by CD (at 222 nm), integrated fluorescence intensity (excited at 280 nm) and fluorescence peak shift are shown in Fig. 3. The fluorescence intensity curve has a very large negative baseline below 40 °C. Thus Trp 86 fluorescence depends strongly on temperature in the native state. A simultaneous fit of all three experimental traces to a two-state model (dotted curves) with arbitrary linear baselines for each trace does not fit the data well. Instead we used a free energy function model to fit the data (solid curves, see below).

On the other hand, Fig. 4 shows that the protein obeys two-state behavior at high concentrations of the denaturant guanidine hydrochloride (GuHCl). A simultaneous two-state model fits that data nicely with a transition midpoint at 1.9 M GuHCl.

Protein folding kinetics

T-jump experiments were carried out at final temperatures from 56 °C to 63 °C, where the largest relaxation signals could be observed. Relaxation kinetics are reported either as the normalized change in tryptophan lifetime (χ_1, see Methods), or as the change in integrated tryptophan fluorescence intensity.

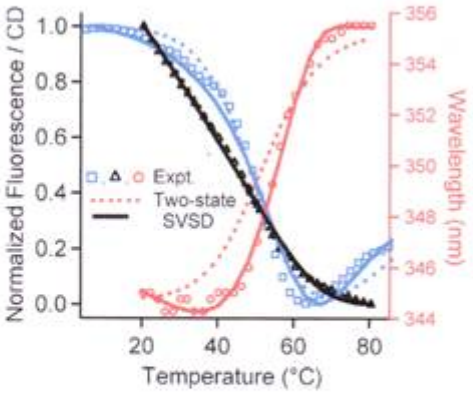


Fig. 3 Circular dichroism (open squares, 3 μM, pH 7) and fluorescence intensity (open triangles, 2 μM, pH 7) thermal titration curves of PTB1 : 4W were measured simultaneously and normalized to the 0 to 1 range for comparison. The fluorescence wavelength shift (open circles) was measured separately on a fluorimeter. No satisfactory global two-state fit was achieved (dashed lines). The SVSD model produced a satisfactory fit (solid lines).

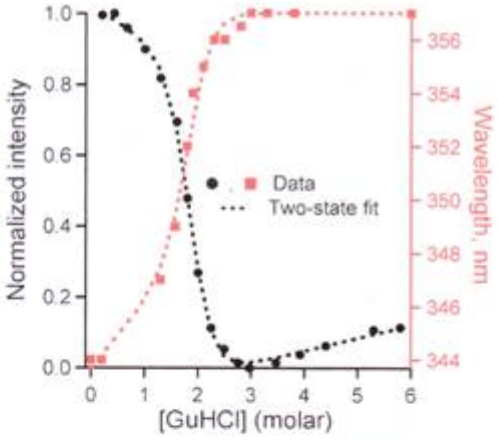


Fig. 4 Fluorescence intensity (circles) and wavelength shift (squares) upon guanidine hydrochloride denaturation of PTB1 : 4W. This data could be fitted by a global two-state model (dashed lines).

Fig. 5 shows the observed relaxation at 60 °C contains a very fast phase of 18 μs, followed by a still fast phase of < 500 μs. The data in Fig. 5 are not compatible with a two-state scenario (single exponential relaxation), but could be fitted by a double-exponential function. The observed slower rate coefficient was temperature-independent within measurement uncertainty over the temperature range we measured.

Stopped-flow experiments were carried out at room temperature (23 °C) to check whether intermediate states accumulate during refolding when the protein is unfolded by GuHCl. Based on the GuHCl titration measurements (Fig. 4), PTB1 : 4W is completely denatured at 3 M GuHCl. Therefore

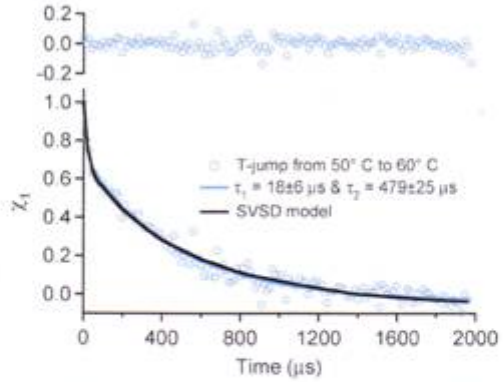


Fig. 5 Folding relaxation kinetics of PTB1 : 4W by a temperature jump from 50 °C to 60 °C detected by normalized fluorescence lifetime change (see Methods). The solid black curve is the SVSD model fit from Smoluchowski dynamics. The top trace shows the residual of a direct double-exponential fit to τ_1 = 18 μs and τ_2 = 479 μs.

protein solution containing 3 M GuHCl was mixed with a buffer solution to several final denaturant concentrations (Fig. 6). The observed relaxation times > 10 ms are considerably slower than the microsecond relaxation rate observed in

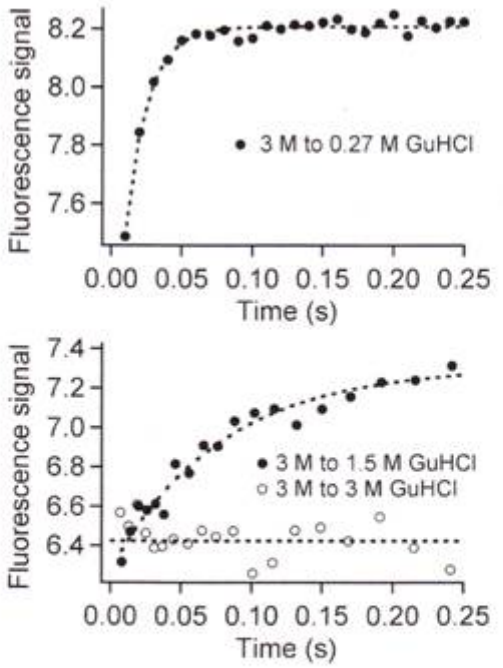


Fig. 6 Stopped-flow experiments at low (0.27 M) and high (1.5 M) final guanidine hydrochloride concentration show slow single exponential refolding kinetics (dashed lines) with no burst phase. For reference, a 3 M to 3 M jump baseline is shown.

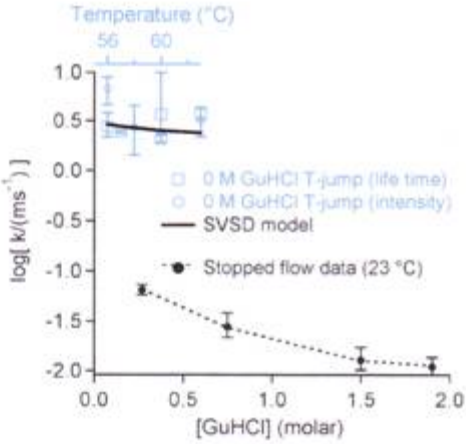


Fig. 7 Summary of the kinetic data. The sub-ms phase upon T-jump is shown in open markers at the top with a temperature axis. The ms refolding kinetics from GuHCl stopped flow are shown in black circles at the bottom. The GuHCl chevron is curved throughout, and the 23 °C stopped flow data do not extrapolate to the much faster T-jump data at 35 °C higher temperature.

0 M GuHCl T-jump experiments (Fig. 5). As seen in Fig. 6, no evidence of a burst phase was found, nor of slower phases up to 20 seconds that exceed 10% of the resolved millisecond phase. The GuHCl refolding transients could be fitted to a two-state (single exponential) model, just like the GuHCl denaturation curves in Fig. 4.

Fig. 7 summarizes the T-jump and stopped flow relaxation kinetics by plotting $\log k = -\log \tau$. Only the slower phase is plotted for the T-jumps. The stopped flow rate at 23 °C does not approach the slower T-jump rates at 0 M GuHCl, and the stopped flow data do not resemble a linear chevron plot. Thus the apparent two-state stopped flow kinetics are not supported by a linear chevron plot.

Model free energy function

The multi-probe thermodynamics and *T*-jump kinetics in Fig. 3 and 5 are not fitted well by a two-state model. Our goal was to determine the simplest one-dimensional free energy function and diffusion coefficient compatible with all the thermal titration and T-jump kinetic data.

The singular value Smoluchowski dynamics (SVSD) model[25] summarized in Fig. 8 uses a genetic algorithm to search the space of folding free energy functions $G(\mathbf{x},T)$, diffusion coefficients $D(\mathbf{x},T)$ and signal functions $S_i(\mathbf{x})$ for the best fit to the data. The functional forms fitted are described in Methods. Our reference coordinate $\mathbf{x}$ was the radius of gyration in nanometers. This choice is arbitrary, but it provides a mapping for the signal functions S onto a reaction coordinate in nm, so one can compare the magnitude of D to known diffusion coefficients (see Methods). SVSD calculates the time-evolving protein population $\rho(\mathbf{x},t)$ after a T-jump, and equilibrium populations $\rho_{eq}(\mathbf{x},T)$ at temperature T, without assigning 'states' and without making the transition state approximation.

We carried out SVSD searches in spaces of 1 and 2 reaction coordinates to globally fit the temperature-dependent data in Fig. 3 and 5. A 1-D surface was sufficient to fit the data. The best 1-D fits to the data are shown as solid lines in Fig. 3, 5 and 7. Fig. 8 shows the corresponding free energy function at three different temperatures. It has three local minima, which are labelled N, U′ and U. The signal functions are also shown in Fig. 8. The genetic algorithm does not guarantee that these free energy and signal functions are unique solutions, but they are representative of the type of functions required to explain the experimental data. We fitted only the average diffusion coefficient of $D = 1.5 \times 10^{-5}$ nm^2 ns^{-1}, independent of position. Different fixed values of D did not produce good fits, but the experimental data is not sufficient to determine the position dependence of $G(\mathbf{x})$ and $D(\mathbf{x})$ independently. All fitting parameters for the best fit are listed in the Electronic Supplementary Information.†

Discussion

PTB1 : 4W has some characteristics of a two-state folder: single exponential relaxation in GuHCl solution and coincident GuHCl denaturation curves. Other characteristics are more representative of a rough free energy surface: double-exponential T-jump kinetics and non-coincident thermal denaturation curves. This discrepancy can be explained if, in the presence of denaturant, a single large barrier partitions the reaction coordinate into a 'folded' and an 'unfolded' basin, whereas in the absence of denaturant, no single barrier dominates. The latter scenario matches the general situation the SVSD model uncovered *via* the genetic algorithm: without denaturant, a folded state and two unfolded states U and U′ with energies within a few RT compete with one another. The barriers are not quite low enough for downhill folding and not quite high enough so that the local minimum U′ can be assigned to a separate folding intermediate. Instead, the observed thermodynamics and kinetics correspond to a hybrid mechanism. This observation is in keeping with the observed T-jump relaxation time, which lies between the few millisecond relaxation time of fast apparent two-state folders and the few microsecond relaxation time of downhill folders.

GuHCl denaturation differs from thermal denaturation. As suggested by Tanford, proteins are more completely unfolded when they are denatured by GuHCl, while proteins under thermal denaturation have some persistent structure that may lower the folding barrier.[26] As discussed by Naganathan *et al.*,[27] denaturant-induced unfolding of small fast-folding proteins at room temperature entails higher activation energies than thermally induced unfolding. Our SVSD surface favors the more structured state U′ over U at high temperature (Fig. 8). We propose that addition of GuHCl favors the less structured state U by raising the U′–U barrier and decreasing the U free energy. Thermally, the protein denatures to U′ rapidly, while in denaturant, it forms U more slowly. We recently observed a candidate for residual structure in U′: 'extended structure' in several proteins at high temperature has more ordered side chains than a random coil, and short segments of beta strand-like geometry.[28]

The switch between two-state kinetics and nearly downhill dynamics is plausible according to the model of Wolynes and

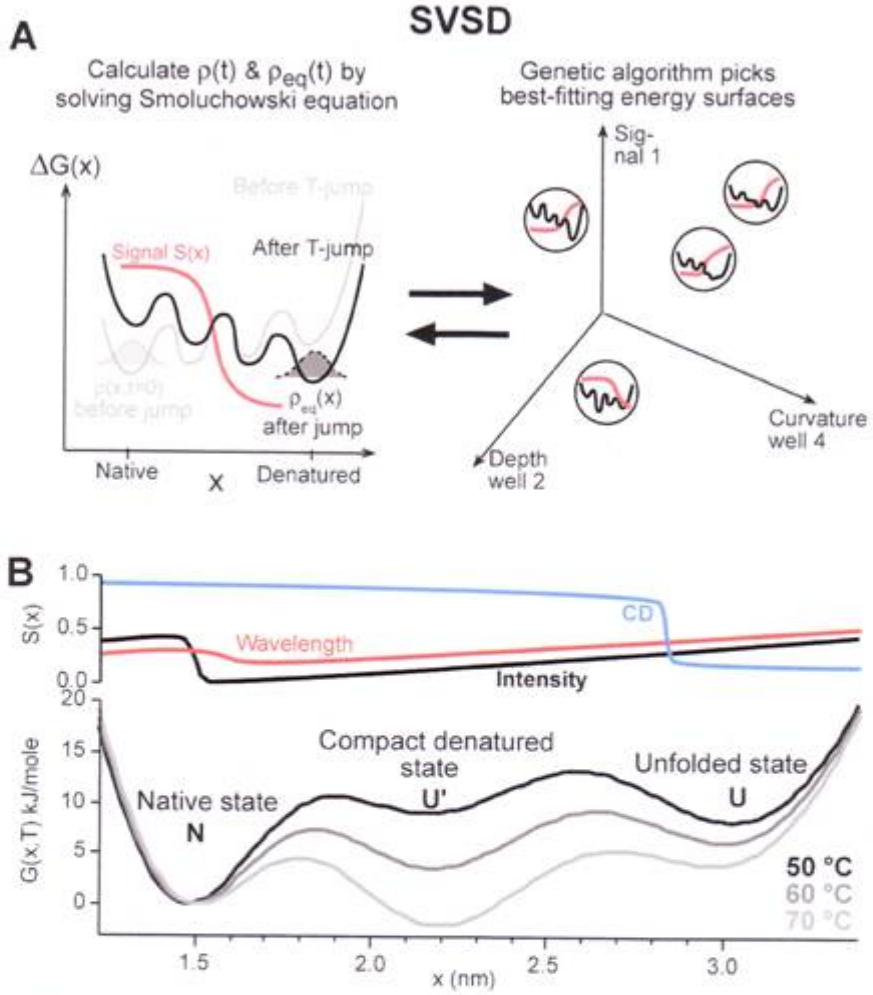


Fig. 8 (A) SVSD method. Left: the protein population distribution ρ is calculated at equilibrium or during kinetics on the free energy surface $G(x)$; then $S(x)\cdot\rho(x)$ is integrated to yield the signal $S(T)$ or $S(t)$. Right: A genetic algorithm selects the best members from a population of free energy surfaces and generates the next generation to calculate signals for comparison with experiment. (B) SVSD free energy functions calculated at 50, 60 and 70 °C. Shown are the optimal functions fitted to the temperature jump and titration data in Fig. 4 and 6. The normalized signal functions are shown at the top.

coworkers.[10] Naganathan *et al.*[27] calculate a denaturant sensitivity of the free energy such that even the slow rate of 12 s^{-1} we observed in the presence of 1.5 M GuHCl makes PTB1 : 4W a fast incipient downhill folder (barrier $< 3RT$) in the absence GuHCl, in accordance with the data in Fig. 7.

In downhill folding, the fastest phase (τ_1 = 18 µs in Fig. 5) is usually labelled "τ_m" for "molecular phase," indicating that it corresponds to the barrier-free ($\Delta G^{\ddagger} \leq RT$) diffusion time across the reaction coordinate. Experiments show that τ_m depends on the topology and the size of the protein. Small downhill folders have τ_m in the range of 0.1 to 2 µs near the thermal denaturation transition,[17,20,29–31] and perhaps an order of magnitude slower at room temperature.[27] One would expect that downhill folders with more complicated topology have slightly longer τ_m. The 415 residue protein phosphoglycerate kinase has $\tau_m \approx 10$ µs under denaturing condition.[32] It is possible that the 91-residue protein PTB1 : 4W with its $\alpha + \beta$ topology has a similarly slow molecular time scale. More likely, and supported by the SVSD fit in Fig. 8, very small residual barriers remain.

SVSD does not allow the arbitrary baselines frequently used for 'two-state' fitting; rather, signal changes arise because the protein distribution shifts along the reaction coordinate, sampling different signals $S(\mathbf{x})$. The large fluorescence intensity baseline below the main unfolding transition (Fig. 3) is caused by a switch of the SVSD signal function. Again, we have

observed a candidate for such intensity changes: tryptophan sidechain fluctuations in the native state that loosen upon increasing temperature can change the fluorescence quantum yield.[33] 'Two-state' fits that require large baselines should be suspect, as these baselines could signal low barriers or large shifts in free energy minima, and hence a breakdown of the two-state approximation.

The actual diffusion coefficient of the PTB1 : 4W folding reaction should be coordinate-dependent,[34–36] but even so, the average value 1.5 x 10^{-5} nm^2 ns^{-1} we fitted is at the low end of values expected for a polypeptide chain. A larger value could not fit the experimental data with a 1-D free energy function. The small diffusion coefficient can be explained in two ways. It could model additional small traps on the free energy landscape, indicating that the true free energy surface of PTB1 : 4W is rougher than the 3 well model shown in Fig. 8. Alternatively, the collective nature of our single reaction coordinate, which reflects diffusion in a multidimensional coordinate space of protein backbone and sidechains, may result in the smaller effective diffusion coefficient in one dimension.[36] A similar result was also reported for folding of the lambda repressor fragment, where a 1-D free energy model required a diffusion coefficient much smaller than the free chain diffusion value of ≈ 0.05 nm^2 ns^{-1}.[23]

The minimal folding barriers in absence of denaturant make PTB1 : 4W an ideal candidate for protein engineering to produce a downhill folder. We make some suggestions of how such a redesign could be achieved. To increase the magnitude of the fluorescence change for kinetic studies and to increase native bias, mutations L68Y or L71Y are good candidates based on the PDB structure 1qm9.[37] In the lambda repressor fragment, an analogous Q33Y mutation places a tyrosine in contact with the fluorescent tryptophan probe, producing a great signal enhancement and stabilizing the native state *via* aromatic stacking interactions.[38,39] To stabilize the secondary structure, the helix propensity of helix 1 can be increased from 4.98% to 21.03% as predicted by AGADIR[40] by replacing Val32 and Ser35 with Ala; turn formation is promoted by shortening the long loops such as the 9-residue loop 5 or the 6-residue loop 1 by inserting four-residue turn motifs.

In summary, PTB1 : 4W is a natural 'missing link' between downhill and activated folding, even without extensive sequence engineering. Its free energy landscape is likely to be rougher (contain more local minima) than the minimal fitting model in Fig. 8; otherwise the fitted diffusion coefficient would have been larger. We suggest mutations that could reduce the ≈ 400 μs 'slow' phase towards the expected molecular time τ_m = 1–10 μs, which the faster phase in Fig. 5 already approaches. Addition of GuHCl considerably reduces the bias towards the native state and increases the folding barrier.

Methods

Protein sample

The RRM4 domain of human PTB1 protein was subcloned by polymerase chain reaction from the human PTB1 gene into the pET28A plasmid (obtained from Professor Doug Black, UCLA). The RRM4 gene fragment with an added NcoI restriction enzyme cutting site at the N-terminus and a HindIII cutting site at the C-terminus was cloned into an Isopropyl-β-D-1-thiogalactopyranoside-inducible (IPTG-inducible) ptac plasmid. The RRM4 gene fragment begins with the Met-Gly at position 440 of full length PTB1 protein, and extends through the natural terminus (isoleucine 531). RRM4W was generated by mutation of Phe526 to Trp using the QuickChange site-directed mutagenesis kit (Stratagene). Both DNA constructs were sequenced for verification.

Protein was expressed in *E. coli* BL-21 DE3 gold cells grown at 37 °C. At OD_{600} = 0.9, 1 mM IPTG was used to induce protein overexpression. Cells were harvested after 4 hours, pelleted by centrifugation at 6500 rpm, washed with buffer A (20 mM tris buffer, pH 7.5, 20 mM NaCl, and 2 mM EDTA), repelleted, and stored at −80 °C overnight. Cells were resuspended in buffer B (20 mM sodium acetate, pH 5.3, 50 mM NaCl, 2 mM EDTA) with 20 μg mL^{-1} of the serine protease inhibitor phenylmethylsulphonyl fluoride (PMSF), 50 units g^{-1} of DNase II, and protease inhibitor cocktail (Sigma), and lysed using a French press. The lysate was spun down at 15000 rpm and dialyzed against 1L of buffer B at 4 °C for 3 hours. The dialysis product was cleared *via* centrifugation and loaded onto a CM-sepharose column equilibrated in buffer C (50 mM Tris, pH 7.5 at room temperature, 10 mM NaCl) at room temperature. After flow-through was discarded, the proteins were eluted with the 10–100 mM NaCl gradient of buffer C. Fractions from the column were combined and concentrated using Vivaspin centrifugal concentrators, and then exchanged into buffer. Freshly prepared 2 mM protein samples in 50 mM sodium phosphate and 150 mM NaCl buffer at pH 7 were stored at 4 °C for further measurements because freezing and thawing protein samples can cause protein aggregation. Protein purity was verified by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS/PAGE).

Thermodynamic measurements

Temperature-dependent fluorescence spectra were collected on a Varian Cary Eclipse fluorimeter. Circular dichroism (CD) spectra and thermal titrations were acquired on a JASCO J-715 spectropolarimeter equipped with a PMT for fluorescence measurements and a Peltier temperature controller. The stored protein sample was diluted with 50 mM sodium phosphate and 150 mM NaCl buffer at pH 7 to reduce its concentration to about 3 μM. The sample was contained in a square quartz cuvette with 1 cm pathlength and covered with mineral oil to prevent evaporation at the higher titration temperatures. The thermal titration was measured from 2 °C to 98 °C in 2 °C steps. Integrated fluorescence data excited at 280 nm was collected simultaneously with the CD signal at 222 nm. Fluorescence peak wavelength was measured on a fluorimeter. The signals $S_i(T)$ were fitted to a global two-state model with linear baselines:

$$S_i(T) = \frac{S_{Ni}(T_m) + m_{Ni}(T - T_m)}{1 + e^{\Delta G_{ND}/RT}} + \frac{(S_{Di}(T_m) + m_{Di}(T - T_m))e^{\Delta G_{ND}/RT}}{1 + e^{\Delta G_{ND}/RT}} \quad (1)$$

where the free energy difference between the native state (N) and the denatured state (D) was fitted by a quadratic polynomial

$$\Delta G_{ND} = G_N - G_D = \Delta G^{(1)}_{ND}(T - T_m) + \Delta G^{(2)}_{ND}(T - T_m)^2 \quad (2)$$

and T_m is the melting temperature. The two state model was insufficient to fit the three traces in Fig. 3 simultaneously.

Kinetic measurements

Stopped flow experiments were performed with an Applied Photophysics SX.18 MV system with 1 : 1 or 10 : 1 mixing at 23 °C. The protein concentration was 1 μM before mixing and the two buffers used were a 3 M GuHCl/50 mM phosphate/150 mM NaCl solution and a 0 M GuHCl/50 mM phosphate/150 mM NaCl solution. Tryptophan was excited at 280 nm and the integrated fluorescence was collected with a 320 nm cutoff filter (WG 320) before the PMT. The time resolution of the collection was 1 ms and the data were acquired with linear time base or logarithm time base for 1 s to 20 s.

Temperature jump-induced relaxation kinetics were measured on our home-built nanosecond laser induced T-jump apparatus.[41] Protein samples were prepared in a 50 mM phosphate and 150 mM NaCl buffer containing 20% D_2O at pH 7.0 (without isotope effect corrections). The concentration of the protein was kept below 21 μM, and the final temperature was kept below 65 °C to avoid aggregation. Full reversibility was confirmed by checking that the steady-state fluorescence lifetimes at 25 °C before and after the T-jump experiment were identical.

The output from a Nd : YAG laser was Raman shifted by 360 psi H_2 gas to produce 1.9 μm wavelength, 10 ns duration T-jump pump pulses. The pump pulse collimated to 2 mm diameter generated an 8–12 °C T-jump in the protein solution within several nanoseconds. The probe beam was a train of 280 nm sub-ps laser pulses 14 ns apart, generated by tripling the output from a mode-locked Ti-sapphire laser. Tryptophan fluorescence was collected with a photomultiplier tube through a Hoya B370 filter to filter out the scattered incident light. In each measurement, a series of tryptophan fluorescence decays digitized with 500 ps time resolution and spaced by 14 ns was collected for a total of 500 μs (limited by the memory of the digitizer used). Folding kinetics longer than 500 μs were obtained by delaying the trigger position for data and obtaining several overlapped traces. Data were binned into several-microsecond intervals to achieve better signal to noise ratio.

When proteins relax towards a larger unfolded population after the T-jump, the fluorescence decay profile evolves in time. The folding kinetics traces were extracted by applying either χ-analysis or plotting the integrated fluorescence intensity traces.[41,42] χ-Analysis fits the change in fluorescence lifetime to a linear combination of two fluorescence decay basis functions, one right after the T-jump, and the other at the end of the data collection window. All the raw data traces obtained by either method (Fig. 4A) could be fitted to exponential decay functions

$$\text{Signal}(t) = y_0 + A_1 e^{-t/\tau_1} + A_2 e^{-t/\tau_2} \quad (3)$$

The Marquard–Levenberg algorithm used for fitting yielded one standard deviation uncertainties for the relaxation times.[43] A nonzero amplitude for more than a single exponential violates the two-state assumption.

SVSD method

In protein folding kinetics, multi-state master equations are usually the fitting model of choice. States are separated by large barriers so protein population can be assigned uniquely to a state for the purposes of thermodynamic fitting; dwell time in such states is long compared to barrier crossing time, so transition state theory can be used. In contrast, downhill folding involves populations that cannot be assigned uniquely to states, and Langevin dynamics[30] or other physics-based[44] models must be used.

Here we present a model that does not assign protein populations to fixed states separated by large barriers. SVSD achieves this by solving the Smoluchowski equation to determine the protein population probability distribution. The ingredients of SVSD are: a free energy surface $G(\mathbf{x},T)$ which depends on reaction coordinates $\mathbf{x}$ and a perturbation (*e.g.* temperature T), signal functions $S(\mathbf{x})$ that correspond to spectroscopic probes, and a diffusion coefficient $D(\mathbf{x})$ that may be coordinate dependent. For fitting, we assumed D to be constant; the free energy is modelled by a sum of Gaussian 'dimples', and the signal functions are modelled by sigmoids, as described in detail in the Electronic Supplementary Information.† For $\mathbf{x}$, we picked the radius of gyration R_g. R_g is 1.5 nm for the native state based on the 1QM9 structure in the protein data bank,[45] analyzed with VMD.[46] For the unfolded state, we used the consensus value $R_g(\text{nm}) = 0.21\,N^{0.6}$ from ref. 47, which yielded 3.1 nm. The choice of $\mathbf{x}$ is arbitrary in the sense that the experimental reaction coordinates are fluorescence and circular dichroism values, which could be mapped onto any reference coordinate. We chose R_g so the order of magnitude of the diffusion coefficient can be compared with literature values.

We have implemented the SVSD method by combining an efficient singular value integrator for the multi-dimensional Smoluchowski equation with a genetic algorithm search through G, S and D to identify free energy surfaces, diffusion coefficients, and signal functions compatible with the data.[25]

SVSD involves five steps. (1) A family of solutions $\{G, S, D\}$ is generated, subject to physically motivated constraints. (2) Equilibrium populations are evaluated at temperatures T

$$\rho_{eq}(\mathbf{x}, T) \sim e^{-G(\mathbf{x},T)/k_B T} \quad (4)$$

for thermodynamic titrations. Time-dependent populations are evaluated after a jump from $T-\Delta T$ to T by solving the Smoluchowski equation

$$\frac{\partial \rho}{\partial t} = \frac{\partial}{\partial \mathbf{x}}\left\{ D(\mathbf{x}) e^{-G(\mathbf{x},T)/k_B T} \frac{\partial}{\partial \mathbf{x}} [e^{+G(\mathbf{x},T)/k_B T} \rho] \right\}$$
$$\rho(t=0) = \rho_{eq}(T - \Delta T) \quad (5)$$
$$\rho(t=\infty) = \rho_{eq}(T)$$

for modelling fast relaxation measurements. (3) Signal functions $S(\mathbf{x})$ are integrated over equilibrium populations $\rho_{eq}(\mathbf{x},T)$ to

yield thermodynamic signals $S(T)$ and over kinetic populations $\rho(\mathbf{x},t)$ to yield kinetic signals $S(t)$. (4) A least squares 'fitness' criterion ranks all family members. (5) The fittest family members $\{G, S, D\}$ are 'mated' by the genetic algorithm to produce diverse offspring, which replace less fit family members in the next iteration. The result is a diverse family of free energy surfaces, signal functions and diffusion coefficients compatible with the data, from which the optimized solution can be picked.

The optimal parameters and range of parameter values are listed in a table in Electronic Supplementary Information.† These values can be used to reproduce the plots in Fig. 8, except that we have shifted the minimum free energy of the native state to $\Delta G = 0$ in Fig. 8, since only the relative free energy of states matters.

Acknowledgements

This work was supported by a grant from the National Science Foundation (MG), MCB 0613643, and by the National Institutes of Health grant GM0772231 (KBH).

References

1 J. Kubelka, J. Hofrichter and W. A. Eaton, *Curr. Opin. Struct. Biol.*, 2004, **14**, 76–88.
2 M. Gruebele, *C. R. Biol.*, 2005, **328**, 701–712.
3 V. Munoz, *Annu. Rev. Biophys. Biomol. Struct.*, 2007, **36**, 395–412.
4 J. Zhang, W. F. Li, J. Wang, M. Qin and W. Wang, *Proteins: Struct., Funct., Bioinf.*, 2008, **72**, 1038–1047.
5 P. L. Freddolino, F. Liu, M. Gruebele and K. Schulten, *Biophys. J.*, 2008, **94**, L75–L77.
6 P. Bruscolini, A. Pelizzola and M. Zamparo, *J. Chem. Phys.*, 2007, **126**, 215103.
7 S. S. Cho, P. Weinkam and P. G. Wolynes, *Proc. Natl. Acad. Sci. U. S. A.*, 2008, **105**, 118–123.
8 N. Ferguson, P. J. Schartau, T. D. Sharpe, S. Sato and A. R. Fersht, *J. Mol. Biol.*, 2004, **344**, 295–301.
9 Y. Ivarsson, C. Travaglini-Allocatelli, M. Brunori and S. Gianni, *Eur. Biophys. J.*, 2008, **37**(6), 721–728.
10 J. D. Bryngelson, J. N. Onuchic, N. D. Socci and P. G. Wolynes, *Proteins: Struct., Funct., Genet.*, 1995, **21**, 167–195.
11 F. Liu and M. Gruebele, *Chem. Phys. Lett.*, 2008, **461**, 1–8.
12 R. Godoy-Ruiz, E. R. Henry, J. Kubelka, J. Hofrichter, V. Munoz, J. M. Sanchez-Ruiz and W. A. Eaton, *J. Phys. Chem. B*, 2008, **112**, 5938–5949.
13 Y. Zhu, D. O. V. Alonso, K. Maki, C.-Y. Huang, S. J. Lahr, V. Daggett, H. Roder, W. F. DeGrado and F. Gai, *Proc. Natl. Acad. Sci. U. S. A.*, 2003, **100**, 15486–15491.
14 A. Fung, P. Li, R. Godoy-Ruiz, J. M. Sanchez-Ruiz and V. Munoz, *J. Am. Chem. Soc.*, 2008, **130**, 7489–7495.
15 F. Liu, D. Du, A. A. Fuller, J. Davoren, P. Wipf, J. Kelly and M. Gruebele, *Proc. Natl. Acad. Sci. U. S. A.*, 2008, **105**, 2369–2374.
16 H. Nguyen, M. Jaeger, J. Kelly and M. Gruebele, *J. Phys. Chem. B*, 2005, **109**, 15182–15186.
17 F. Liu and M. Gruebele, *J. Chem. Phys.*, 2009, **131**, 195101.
18 F. Liu, C. Dumont, Y. J. Zhu, W. F. DeGrado, F. Gai and M. Gruebele, *J. Chem. Phys.*, 2009, **130**, 061101.
19 W. Y. Yang and M. Gruebele, *J. Am. Chem. Soc.*, 2004, **126**, 7758–7759.
20 F. Liu and M. Gruebele, *J. Mol. Biol.*, 2007, **370**, 574–584.
21 C. Clerte and K. B. Hall, *RNA*, 2006, **12**, 457–475.
22 K. Maki, H. Cheng, D. A. Dolgikh and H. Roder, *J. Mol. Biol.*, 2007, **368**, 244–255.
23 W. Y. Yang and M. Gruebele, *Nature*, 2003, **423**, 193–197.
24 W. Yang, E. Larios and M. Gruebele, *J. Am. Chem. Soc.*, 2003, **125**, 16220–16227.
25 G. Scott and M. Gruebele, *J. Comput. Chem.*, 2009, in press.
26 C. Tanford, *Adv. Protein Chem.*, 1968, **23**, 121–282.
27 A. N. Naganathan, U. Doshi and V. Munoz, *J. Am. Chem. Soc.*, 2007, **129**, 5673–5682.
28 W. Y. Yang, E. Larios and M. Gruebele, *J. Am. Chem. Soc.*, 2003, **125**, 16220–16227.
29 J. Kubelka, W. A. Eaton and J. Hofrichter, *J. Mol. Biol.*, 2003, **329**, 625–630.
30 W. Yang and M. Gruebele, *Biophys. J.*, 2004, **87**, 596–608.
31 D. M. Vu, E. S. Peterson and B. B. Dyer, *J. Am. Chem. Soc.*, 2004, **126**, 6546–6547.
32 J. Sabelko, J. Ervin and M. Gruebele, *Proc. Natl. Acad. Sci. U. S. A.*, 1999, **96**, 6031–6036.
33 J. Ervin, E. Larios, S. Osvath, K. Schulten and M. Gruebele, *Biophys. J.*, 2002, **83**, 473–483.
34 J. Chahine, R. J. Oliveira, V. B. P. Leite and J. Wang, *Proc. Natl. Acad. Sci. U. S. A.*, 2007, **104**, 14646–14651.
35 C. L. Lee, G. Stell and J. Wang, *J. Chem. Phys.*, 2003, **118**, 959–968.
36 R. B. Best and G. Hummer, *Proc. Natl. Acad. Sci. U. S. A.*, 2010, **107**, 1088–1093.
37 M. R. Conte, T. Grune, J. Ghuman, G. Kelly, A. Ladas, S. Matthews and S. Curry, *EMBO J.*, 2000, **19**, 3132–3141.
38 S. Ghaemmaghami, J. M. Word, R. E. Burton, J. S. Richardson and T. G. Oas, *Biochemistry*, 1998, **37**, 9179–9185.
39 W. Y. Yang and M. Gruebele, *Biochemistry*, 2004, **43**, 13018–13025.
40 E. Lacroix, A. R. Viguera and L. Serrano, *J. Mol. Biol.*, 1998, **284**, 173–191.
41 R. M. Ballew, J. Sabelko, C. Reiner and M. Gruebele, *Rev. Sci. Instrum.*, 1996, **67**, 3694–3699.
42 J. Ervin, J. Sabelko and M. Gruebele, *J. Photochem. Photobiol., B*, 2000, **54**, 1–15.
43 W. H. Press, S. A. Teukolsky, W. T. Vetterling and B. P. Flannery, *Numerical Recipes in Fortran: The Art of Scientific Computing*, Cambridge University Press, Cambridge, 2nd edn, 1992.
44 J. Kubelka, E. R. Henry, T. Cellmer, J. Hofrichter and W. A. Eaton, *Proc. Natl. Acad. Sci. U. S. A.*, 2008, **105**, 18655–18662.
45 H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov and P. E. Bourne, *Nucleic Acids Res.*, 2000, **28**, 235–242.
46 W. F. Humphrey, A. Dalke and K. Schulten, *J. Mol. Graphics*, 1996, **14**, 33–38.
47 N. C. Fitzkee and G. D. Rose, *Proc. Natl. Acad. Sci. U. S. A.*, 2004, **101**, 12497–12502.

Chapter 8

Dynamic coupling of PTB RRM3 and RRM4

One of the more interesting implications of the interaction between PTB RRM3 and RRM4 is the possibility of allosteric connection of the binding sites in PTB1:34. We already know that the interaction changes both the motional properties of the protein and its ability to bind RNA. However, the root of these changes remains to be determined. The fact that RRM4 is folded and quite stable as an individual domain, yet only participates in the RNA binding event when coupled to RRM3 in the context of PTB1:34 suggests some form of communication between the binding sites of RRM3 and RRM4 (Maynard and Hall, 2009). Since no large scale structural rearrangements occur as a result of the interaction, the means of communication must be more subtle, possibly involving dynamic, small scale structural changes.

Minor rearrangements of secondary structure usually occur on a microsecond timescale. This leads to the hypothesis that the communication between PTB RRM3 and RRM4 binding sites may occur via a network of non-covalent contacts that are ushered through the protein by small scale rearrangements of secondary structure. This idea is consistent with the increased motions and decreased secondary structure observed upon interaction of these RRMs. Guided by a recently developed method to identify pathways of non-covalent contacts (Bradley et al., 2008), we have produced a mutant protein that is likely to decouple the RNA binding sites without physically separating the domains. This mutant protein will allow us to probe further into the importance of communication between the RRMs, particularly in regard to RNA binding and protein dynamics.

Identification of communication pathways

Possible paths between the binding sites of RRM3 and RRM4 were determined using the method described in Bradley et al., 2008. Binding sites were defined as the set of residues involved in RNA binding of each domain in PTB1:34 for both RNA tetramer binding studies (Oberstrass et al, 2005) and the binding studies involving a longer RNA described in Chapter two. Using numbering from PDB ID:2EVZ (Oberstrass et al., 2005), these residues include: G12, V15, V16, N22, Y38, V41, Q42, R43, I83, K105, I126, H134, L135, K169, L199, V201, S202, F203, S204, K205 and I208. It is important to point out that the method for determining non-covalent contacts does not account for hydrophobic contacts. Since the interaction between the RRMs is primarily hydrophobic (Vitali et al., 2006), a good deal of the possible pathways may be missed. However, any pathways that do emerge from this analysis will be likely to include residues that can be mutated with less chance of protein destabilization.

The above analysis yielded a single residue, arginine 114 (numbering from PDB ID: 2EVZ), that was present in 80% of all possible paths between binding sites. Likely there exist many more paths that were missed due to incomplete binding site definitions (as discussed in chapter 2), as well as the fact that the analysis does not account for hydrophobic interactions. Nonetheless, the goal was to identify a single residue for mutation, so a mutant with R114 changed to an alanine was prepared (Figure 1).

The PTB1:34R114A mutant

The DNA for the PTB1:34R114A mutant was generated using a Stratagene site

directed mutagenesis kit. PTB1:34 DNA was used as a template, and the primers were 5'-CTCACCCCTGCACGCCTTCAAGAAGCCG-3' (top) and 5'-CGGCTTCTTGAACGGGTGCAGGGGTGA-3' (bottom). Mutated DNA was transformed into E. coli BL-21 DE3-gold competent cells, and protein production and purification ensued exactly as for PTB1:34 (chapter 2).

Initial characterization began by comparison of secondary structure using CD, and shows that the spectra of PTB1:34 and PTB1:34R114A are nearly indistinguishable, as expected (Figure 2). GndHCl denaturation revealed that the proteins have similar stability, but slightly difference profiles. In particular, the initial increase in ellipticity observed at low concentrations of denaturant in PTB1:34, that was tentatively attributed to domain separation, is missing in the denaturation profile of the mutant (Figure 2). This result is encouraging as it shows that the thermodynamic properties of the mutant are indeed different from the wild type protein, however, more work needs to be done to ascertain whether or not the domains are uncoupled but not physically separated. Simple RNA binding studies are also necessary to determine the functional consequences of the mutation.

Initial attempts to prepare a ^{15}N labeled sample for NMR backbone comparisons were not successful. While the protein expresses well in LB media, the expression is poor in M9 minimal media, even though cell growth does not seem to be compromised. However, after only two attempts, there is no reason to think that this obstacle is insurmountable. In the meantime, many RNA binding studies could be done that would

be very informative.

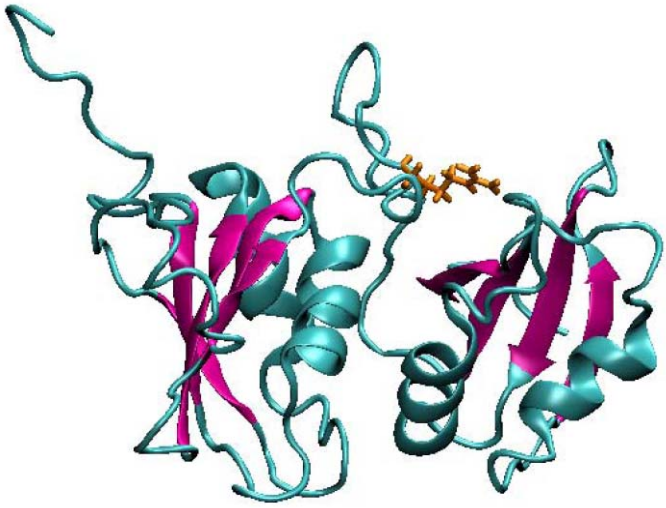
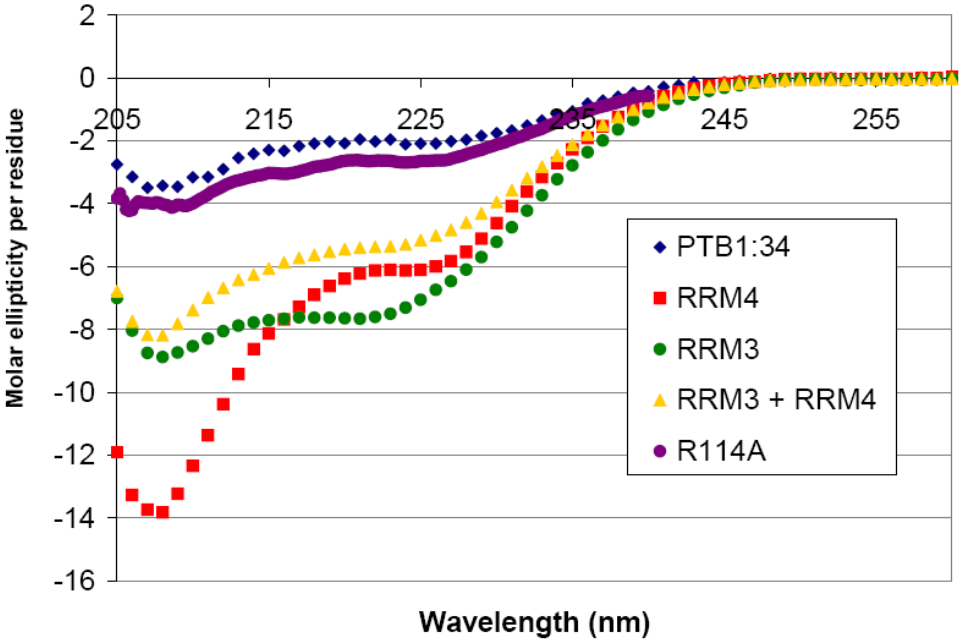


Figure 1: Arginine 114 may be important for communication between PTB RRM3 and RRM4. The three-dimensional structure of PTB1:34 is shown (PDB ID:2EVZ) with the β-sheet surfaces in magenta (the putative RNA binding surfaces) and R114 in orange licorice. R114 was mutated to an alanine in an attempt to decouple the domains without physical separation.


2
0
-2
-4
-6
-8
-10
-12
-14
-16
205
215
225
235
245
255
Molar ellipticity per residue
Wavelength (nm)
PTB1:34
RRM4
RRM3
RRM3 + RRM4
R114A

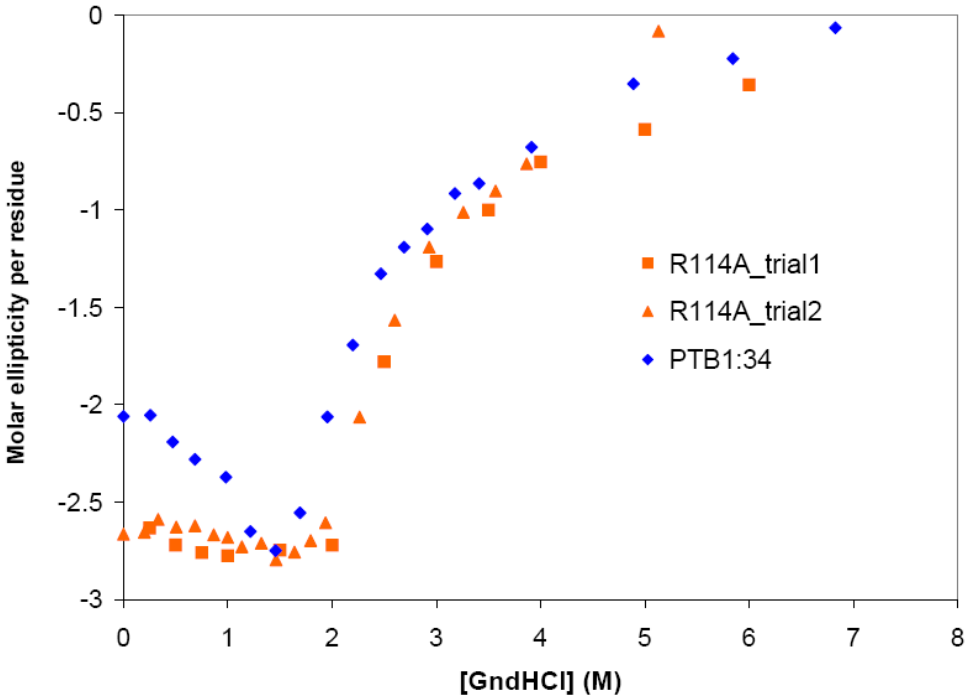

0
-0.5
-1
-1.5
-2
-2.5
-3
0
1
2
3
4
5
6
7
8
Molar ellipticity per residue
[GndHCl] (M)
R114A_trial1
R114A_trial2
PTB1:34

Figure 2: The secondary structure of PTB1:34 is perturbed by the R114A mutation. CD spectra of all protein constructs used in this work is shown in the top panel with PTB1:34 in blue and PTB1:34R114A in purple. Net secondary structure is gained with the mutation. In addition, the chemical denaturation of R114A is significantly different from PTB1:34 (bottom panel). Importantly, the initial increase in CD signal observed in PTB1:34 (blue) that was tentatively attributed to separation of the domains is missing in the mutated protein (orange). These results suggest that the mutation may indeed decouple the two RRMs. However, structural studies will be necessary to confirm this, as well as to make sure that the domains in the mutant are not physically separated.

References

Bradley, M. J., Chivers, P. T. and Baker, N. A. Molecular dynamics simulation of the *Escherichia coli* NikR protein: Equilibrium conformational fluctuations reveal interdomain allosteric communication pathways. (2008). *J. Mol. Biol.*, **378**, 1155-1173.

Maynard, C. M. and Hall, K. B. Interactions between PTB RRMs induce slow motions and increase RNA binding affinity. (2009) *J. Mol. Biol.*, **397**, 260-277.

Oberstrass, F. C., Auweter, S. D., Erat, M., Hargous, Y., Henning, A., Wenter, P., Reymond, L., Amir-Ahmady, B., Pitsch, S., Black, D. L. and Allain, F. H.-T. Structure of PTB bound to RNA: Specific binding and implications for splicing regulation. (2005) *Science*, **309**, 2054-2057.

Vitali, F., Henning A., Oberstrass, F. C., Hargous, Y., Auweter, S. D., Erat, M. and Allain, F. H.-T. Structure of the two most C-terminal RNA recognition of PTB using segmental isotope labeling. (2006) *EMBO J.*, **25**, 150-162.

Concluding Remarks

Unique properties of the two C-terminal RRMs of PTB create a system that, although challenging, is interesting to study on many levels. PTB has been found to participate in a wide range of cellular tasks, though it remains unclear how it does so. To this end, the thorough investigation of protein structure, unfolding properties and protein dynamics presented in this work will hopefully contribute to a greater understanding of how this versatile protein functions.

The fact that the protein consists of RRMS, protein motifs that have been exhaustively studied in terms of structure and binding, enables more general questions about structure function relationships to be asked. Particularly, the C-terminal extension in RRM3 provides a tool that can be exploited to gain insight into how structural modifications of a common motif may be used by nature to alter protein binding events.

Attempts to characterize the thermodynamic stability of the proteins used in this study revealed a linear denaturation profile for RRM4. This discovery was used to learn more about protein unfolding energy landscapes, and RRM4 was classified as a natural missing link between downhill and activated unfolding scenarios. In addition, since an array of states are potentially accessible along the unfolding pathway, further study of this RRM may be useful for understanding structural and thermodynamic changes that occur as proteins unfold.

Finally, the nature of the interaction between PTB RRM3 and RRM4 implies allostery. We have designed a protein mutant intended to disrupt the communication between the two RRMs without physically separating them. This protein expresses well and is stable, facilitating future structure and dynamics studies that may be of great interest.

Hopefully, the work presented herein not only enhances our understanding of protein systems but opens the door for future studies as well.